Vol. 50 APRIL 1959 No. 2

The Journal of Educational Psychology

CONTENTS

Level of Aspiration and Academic Success 47
LEONARD WORELL

Some Characteristics of High School Pupils From Three Income Groups 55
JOHN K. COSTER

Measurement Characteristics of Recall in Relation to the Presentation of Increasingly Large Amounts of Material 63
WILSE B. WEBB AND MARVIN SCHWARTZ

Personality Syndromes and Academic Achievement 66
GEORGE MIDDLETON, JR. AND GEORGE M. GUTHRIE

Predicting Principals' Ratings of Teacher Performance From Personality Data 70
ROBERT F. PECK

The Relationship of Home and School Experiences to Scores on Achievement Tests 75
JOHN W. FRENCH

The Effect of an Introductory Psychology Course on Self-Insight 83
FRANK COSTIN

Retention in Arithmetic Among Children of Low, Average, and High Intelligence at 117 Months of Age 88
HERBERT J. KLAUSMEIER AND JOHN F. FELDHUSEN

AMERICAN PSYCHOLOGICAL ASSOCIATION

The Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

LEE J. CRONBACH
University of Illinois
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
HAROLD E. JONES
University of California
IRVING LORGE
Teachers College, Columbia University
DOROTHEA McCARTHY
Fordham University
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
PERCIVAL M. SYMONDS
Teachers College, Columbia University
GEORGE G. THOMPSON
Syracuse University

The Journal of Educational Psychology publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

SARAH WOMACK
Editorial Assistant

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Mt. Royal & Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Baltimore, Md.
Copyright, 1959 by the American Psychological Association, Inc.

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 50 April, 1959 Number 2

LEVEL OF ASPIRATION AND ACADEMIC SUCCESS[1]

LEONARD WORELL

State University of Iowa

Interest in the predictive value of "nonintellectual" factors in academic achievement has recently increased as a consequence, in part, of the limited predictive value of conventional academic aptitude measures. Only a small portion of the variance in academic success can be attributed to the variance in measures of ability and previous academic achievement. Unfortunately, the findings of available studies dealing with "nonintellectual" variables do not seem to contribute appreciably to either the practical aim of improved prediction or the theoretical purpose of increasing the generality of behavior theories.

The present study employs the level of aspiration paradigm in predicting academic achievement. Most commonly, the level of aspiration refers to the subsequent level of performance which an individual states he anticipates achieving following performance on a task. Major emphasis is on the discrepancy between previous performance and expected subsequent performance. It has been apparent for some time, however, that the level of aspiration is not a unitary construct. The use of a number of levels of aspiration is exemplified in the work of Lewin, Dembo, Festinger, and Sears (1944), where the levels a person *hoped for, expected,* and was *minimally satisfied with* in achieving were viewed as separately meaningful. The level of aspiration, therefore, may be conceptualized more broadly and a variety of discrepancy scores may be obtained depending upon the nature of the questions which elicit aspiration statements. Not only may discrepancy scores be derived from expectancy estimates and previous performance, but also from a comparison of wished for performance and previous performance, wished for performance and expected performance, and so on. In the present study, discrepancy scores are obtained by using various *estimates* of an individual's own performance rather than his actual performance.

Discrepancy scores derived from estimates of performance may also be viewed as reflecting a dimension of reality-irreality. The relationship of level of aspiration to a reality-irreality dimension has appeared in the work of Lewin et al. (1944) and Rotter (1954). In this study, the reality-irreality dimension is defined as any discrepancy between estimated performance or effort and some other aspiration estimate. The larger a discrepancy the more an individual may be regarded as unrealistic.

The general hypothesis of the study is that the reality-irreality dimension is related to academic success. It is assumed that persons with highly discrepant scores base their estimates of performance on unrealistic considerations of a wishful or

[1] This research was supported in part by a grant-in-aid from the Social Science Research Council. The author wishes to thank the members of the Reed College faculty for their cooperation in this project. Particular appreciation is extended to Fred Courts, under whose direction the questionnaire employed in the present study was developed.

avoidant nature. Such individuals, when faced with academic performance situations, are expected to invoke more unrealistic and avoidant responses than *Ss* with lower discrepancy scores. Thus, for example, of two persons with identical estimates of previous performance but divergent estimates of subsequent performance, the one with the more discrepant score is expected to perform more poorly since achievement situations for him evoke more unrealistic behaviors. It should be noted that we are not concerned here with the factors leading to the development of unrealistic behavior, but only with the relationship of this behavior to academic performance.

Experimental Hypotheses

Level of aspiration measures were constructed on the basis of the following questions, which were rated by all *Ss* on a decile scale:

1. How hard do you work on your studies relative to other students?
2. How do you think your average grades compare with those of your classmates?
3. If you plan to return next year, how well do you expect to do in comparison with other members of your class?
4. If you really tried to do well and worked near the limits of your capacity, how would your average grades compare with those of your classmates?
5. How well would you like to do in order to be reasonably well satisfied according to your own standards?

The following predictions refer to the combination of *pairs* of the preceding estimates.

Academic adjustment is inversely related to the degree to which students' estimates of:

1. What their performance would be when working near the limits of capacity exceed their estimates of how hard they have worked in the past (Questions 4 minus 1).
2. Their future performance are above their estimates of previous performance (Questions 3 minus 2).
3. What their performance would be when working near the limits of capacity are above their estimates of previous performance (Questions 4 minus 2).
4. What they would be reasonably satisfied with in their performance exceeds their estimates of previous performance (Questions 5 minus 2).

Method

Subjects

The study was conducted in a small liberal arts college which places a strong emphasis on scholastic achievement. The college attempts, however, to encourage a desire for learning rather than emphasize grade point achievement. Thus, grades are assigned for administrative purposes but are not typically revealed to the student by the instructors. The student is given only an indication of his relative decile standing at the end of each year. Students in the freshmen class, in contrast to the other classes, had been given no indication of their relative positions, since the study was conducted prior to the end of their first year.

The 421 students used in this study were drawn from the entire student body of about 550 for the academic year 1952–53. Subjects participated on a voluntary basis. The ratio of males to females in both the sample and the total enrollment is approximately four to one.

Materials

Two major types of data were obtained: level of aspiration data and criterion measures: the latter included (*a*) total college grade average, computed by placing each student into one of three categories, above C, C, and below C, for each year of his attendance, and then averaging across years; (*b*) grade decile rank during the year that the information was obtained; (*c*) survival, attainment or nonattainment of a degree.

TABLE 1

RELATIONSHIP OF APTITUDE AND HIGH SCHOOL ACHIEVEMENT TO DECILE RANK AND TOTAL GRADE AVERAGE

	Freshmen (N = 138)				Sophomores (N = 97)			
	CEEB	HSA	DR	TGA	ACE	HSA	DR	TGA
Aptitude	—	.35*	.34*	.30*	—	.13	.17	.19
High School Achievement		—	.53*	.44*		—	.31*	.41*
	Juniors (N = 76)				Seniors (N = 78)			
	ACE	HSA	DR	TGA	ACE	HSA	DR	TGA
Aptitude	—	.13	.15	.13	—	.15	.13	.19
High School Achievement		—	.34*	.23*		—	.23**	.36*

* Significant at .01 level.
** Significant at .05 level.

RESULTS AND DISCUSSION

Efficiency of Traditional Measures

Many studies (Henry, 1950) support the utility of aptitude and high school achievement measures in predicting college success. This study used the ACE with sophomores, juniors, and seniors, and the CEEB with freshmen. Table 1 shows that the CEEB, but not the ACE, correlates significantly with grade decile rank and total grade average. In addition, high school achievement is clearly superior to aptitude in prediction and little (Table 2) is gained by combining high school achievement and aptitude.

TABLE 2

MULTIPLE CORRELATIONS OF APTITUDE AND HIGH SCHOOL ACHIEVEMENT TO DECILE RANK AND TOTAL GRADE AVERAGE

Class	DR	TGA
Freshmen	.55	.47
Sophomores	.34	.43
Juniors	.35	.23
Seniors	.23	.39

Aspiration and Scholastic Success

Prior to an examination of the results for the specific predictions, it may be worthwhile to determine the stability of the five aspiration questions. Previous studies (Lewin et al., 1944), employing relatively short retest intervals, suggest that level of aspiration is a reasonably consistent measure. The stability of our aspiration measures over a four-year period on a group of 40 is presented in Table 3. Considering the number of factors which may potentially contribute to changes in estimates, over a relatively long time span, the consistency of estimates is quite high.

The following four predictions were examined: Academic adjustment is inversely related to the degree to which students' estimates of (*a*) what their performance would be when working near the limits of capacity exceed their estimates of how hard they have worked in the past; (*b*) their future performance are above their estimates of previous performance; (*c*) what

TABLE 3

STABILITY OF ASPIRATION ESTIMATES OVER FOUR-YEAR PERIOD

Estimate	r^*
1. Effort	.54
2. Previous Academic Performance	.61
3. Expected Performance Next Year	.56
4. Capacity Performance	.52
5. Reasonable Satisfaction Level	.48

* .39 required for significance at .01 level.

their performance would be when working near the limits of capacity are above their estimates of previous performance; (*d*) what they would be reasonably satisfied with in their performance exceeds their estimates of previous performance.

With the exception of the nonsignificant relationship to decile rank in the senior group in Prediction *a*, the results (Table 4) show significant relationships between aspiration discrepancy scores and both decile rank and total grade average in all other comparisons. We may conclude that the student who behaves unrealistically, or more specifically, the one who perceives his reasonable level of performance satisfaction as lying above previous achievement, has aspirations markedly beyond past performance, estimates his potential capacity performance as lying far above the effort he expends, and believes that he can achieve far beyond what he already has by pressing himself to the limit of his ability, will tend to attain and continue to attain a lower scholastic standing. On the other hand, the student who holds moderate aspirations, perceives his effort as being commensurate with his potential capacity performance, does not see his performance as markedly improving by making a "total push," and whose standards of acceptable satisfaction are below his previous achievement, tends to obtain grade success.

Note may be taken of the generally declining values of the correlations with grade average from the sophomore to the senior

TABLE 4

INTERCORRELATIONS AMONG PREDICTOR VARIABLES AND THE RELATIONSHIP OF THE PREDICTORS TO DECILE RANK AND TOTAL GRADE AVERAGE[a]

1	2	3	4	5	6	DR	TGA	1	2	3	4	5	6	DR	TGA
Freshmen (N=138)								Juniors (N = 85)							
1	.32	.67	.23	.35	−.16	−.22**	−.26**	1	.49	.68	.19	.17	−.24	−.25*	−.22*
2		.73	.71	−.06	−.32	−.56**	−.48**	2		.69	.45	.19	−.21	−.28**	−.42**
3			.66	.01	−.40	−.57**	−.63**	3			.64	.04	−.17	−.53**	−.57**
4				.26	−.36	−.65**	−.58**	4				−.13	.14	−.22*	−.51**
Sophomores (N = 99)								Seniors (N = 75)							
1	.40	.69	.17	.14	−.22	−.32**	−.27**	1	—	.30	.19	−.36	.06	−.09	−.24*
2		.69	.71	.38	−.14	−.52**	−.45**	2		—	—	—	—	—	—
3			.65	−.17	−.17	−.60**	−.70**	3			.77	.04	−.28	−.48**	−.35**
4				−.15	−.24	−.57**	−.58**	4				.14	−.34	−.38**	−.26**

[a] Variables include: 1. Estimated capacity performance minus estimated effort. 2. Estimated expected performance minus estimated previous performance. 3. Estimated capacity performance minus estimated previous performance. 4. Estimated reasonable satisfaction level minus estimated previous performance. 5. Aptitude. 6. High school achievement.

* Significant at .05 level.

** Significant at .01 level.

years. Tentatively, at least two factors may be involved in this progression: either those with highly discrepant scores have discontinued their education, thus making each successive group more homogeneous, or some modification may occur in the apparent disabling aspects of high standards. An examination of the scores obtained by the four classes indicates that they do not differ with regard to means and variances of discrepancy scores. It would appear, then, that although highly discrepant scores are associated with poor early attainment, some students are capable of learning to adjust to unrealistic standards and to function adequately.

The high degree of relationship apparent among the predictors (Table 4) may be attributable to the common underlying dimension of reality-irreality. On the other hand, these relationships may also point to the tendency of an individual to rate himself consistently in a given direction and may involve response sets analagous to those proposed by Cronbach (1950).

Multiple Correlations

Table 5 contains the multiple correlations for ability and high school achievement, aspiration indices, and the combined indices with both decile rank and total grade average. The use of all six variables provides the highest multiple correlations with both academic criteria. Although somewhat lower than those obtained with six variables, the multiple *R*s with the four aspiration discrepancy measures exceed those obtained with the ability and high school achievement indices. The latter correlations strongly support the value of these relatively nonintellectual indices in predicting academic success.

Survival

In approaching the problem of mortality, the aspiration discrepancy scores were combined into a single score (AS) for each student. This procedure seemed justified since the correlation of the four additively combined aspiration scores with total grade average yielded values of $-.57$ and $-.54$ for the freshmen and sophomore classes respectively, both of which are significant beyond the .01 level. Using the AS, an empirical cutting point was determined for discriminating between graduates and nongraduates. By adopting a cutting score of 9 and below versus above 9, a chi square of 16.14, significant beyond the .001 level, was obtained for the freshmen group. It is apparent from Table 6 that the AS is not equally discriminatory for both high and low scorers. Freshmen high scorers are differentiated into graduates and nongraduates with reasonable accuracy while low scorers are not. With the sophomore group, a chi square of 7.64 is obtained, significant at beyond the .01 level. In contrast to the freshmen, sophomore high scorers are not as clearly differentiated into graduates and nongraduates, while low scorers are more satisfactorily separated. Finally, Table 6 indicates that the AS is generally less satisfactory for the juniors. The chi square is reduced to 6.08, significant at less than the .02 level. Furthermore, in contrast to the freshmen and sophomore groups, high scorers are not satisfactorily separated into

TABLE 5

MULTIPLE CORRELATIONS OF ASPIRATION AND INTELLECTUAL MEASURES TO DECILE RANK AND TOTAL GRADE AVERAGE

Variables[a]	Freshmen		Sophomores		Juniors		Seniors	
	DR	TGA	DR	TGA	DR	TGA	DR	TGA
5, 6	.55	.47	.34	.43	.35	.23	.23	.39
1, 2, 3, 4	.68	.69	.66	.79	.75	.63	.48	.39
1, 2, 3, 4, 5, 6	.75	.72	.74	.85	.70	.66	.50	.52

[a] The correlations at the senior level do not include Variable 2. For key to numbered Variables 1-6 see bottom of Table 4.

TABLE 6
FREQUENCIES OF FRESHMEN, SOPHOMORE AND JUNIOR GRADUATES AND NONGRADUATES AND WITH HIGH AND LOW AS SCORES

	AS Scores	
	>9	9 and <
Freshmen		
Graduates	7	36
Nongraduates	41	32
Sophomores		
Graduates	11	35
Nongraduates	24	20
Juniors		
Graduates	12	46
Nongraduates	15	16

graduates and nongraduates, whereas low scorers are differentiated more accurately.

Aspiration, Grades, and Attrition

It may reasonably be asked what relationship grades have to both the AS and attrition. Table 7 shows that among below C students a significantly greater number drop out than remain to receive a degree (50 vs. 11), while the reverse is found among C and above C students (60 vs. 94). Though grades are an obvious factor in mortality, these data indicate the presence of a large group with grades of C and above who leave for other reasons.

Since the greatest attrition occurs during the freshmen and sophomore years (Research Report, 1953), these classes have been combined to increase our totals. The portion of Table 7 dealing with *below C* students indicates that, despite the relatively large number of high scoring nongraduates, the AS fails to differentiate graduates from nongraduates at a significant level. On the other hand, students with grades of C and above are successfully differentiated into graduates and nongraduates. The chi square of 14.35 is significant beyond the .001 level. Thus, although grades are significantly related to attrition, the AS makes a preliminary contribution toward discriminating graduates from nongraduates, in a group which is not experiencing scholastic deficiency.

While the aspiration measures have been shown to be related to academic success, one factor which might enhance the obtained relationships should be considered. This involves the association between some of the discrepancy aspiration scores and previous performance. A number of studies have found that goal discrepancy scores are to some extent negatively related to the performances from which they are derived. This stems primarily from two factors: (*a*) the greater opportunity for individuals low in performance to set higher goals, thereby obtaining larger discrepancy scores; and (*b*) forcing individuals at the extremes to set goals within a restricted range or in a given direction. Thus, those with high performances would be restricted in their estimates at the upper limit, while those with

TABLE 7
FREQUENCIES OF GRADUATES AND NONGRADUATES AMONG STUDENTS DIFFERENTIATED BY GRADES AND AS SCORES

	AS Scores	
	>10	10 and <
Below C Students[a]		
Graduate	5	6
Nongraduate	10	40
C and Above C Students[b]		
Graduate	78	16
Nongraduate	32	28

[a] Below C Students $X^2 = 1.93$ (NS).
[b] C and Above Students $X^2 = 14.35$ (.001).

very low performances would have little choice but to set goals at or above previously attained levels. These conditions leave open the possibility that the discrepancy scores predict well because of their dependence on previous performance. A similar argument has been advanced by Schultz and Ricciuti (1954).

There are two aspects of this study, however, which suggest that this interpretation of our findings is only partially and minimally correct. First, if including previous performance in one of the criteria (total grade average) were influential, one would expect that the correlations between aspiration measures and total grade average would increase from the freshmen to the senior classes. This is apparent since the more advanced students would have more of their previous performance included in the criterion. This expectation is not born out by the results (Table 5), in that the sophomore correlations are the highest, followed by the freshmen and juniors, with the seniors showing the lowest correlations. Thus, the magnitudes of the relationships are almost the reverse of expectation for the respective classes. The second consideration is that the success of the aspiration measures with regard to attrition among relatively successful students (C and above C total average grades) suggests that factors other than previous performance are being assessed by these measures.

Since a number of studies (Gould & Kaplan, 1940; Holt, 1946; Schultz & Ricciuti, 1954) have been unsuccessful in relating level of aspiration measures to academic success, it is worthwhile to note some of the differences between the present study and former ones. At least two differences may be important. The first of these is the type of level of aspiration task employed; in contrast to the use of somewhat artificial tasks, the present study obtained estimates that were highly related to the type of performance which we wished to predict. A second possible factor is that contrary to customary procedure, we did not use the actual previous performance level of the student but rather the student's *estimates* of his previous performance, in a setting where students maintain only a very general impression of their grade standing. This procedure may increase the probability of reality distortion. The further utility of this procedure will have to be evaluated by subsequent research.

Summary and Conclusions

This study attempted to determine the theoretical and empirical utility of the level of aspiration method in predicting college grades and attrition. The general assumption was that discrepancy scores between various estimates related to academic performance reflected the reality-irreality level at which the individual operated. Individuals with large discrepancy scores were considered unrealistic and were expected to perform more poorly in academic situations since they would be more likely to employ unrealistic and avoidant problem solution behaviors.

Discrepancy measures related to the general hypothesis were obtained from the combination of estimates given to varying pairs of five level of aspiration questions. The college sample, consisting of almost the entire student body of a small liberal arts college, estimated their position on a decile scale according to how hard they worked, previous performance, expected future performance, performance when working near the limits of capacity, and level of reasonable satisfaction.

The results suggest the following conclusions:

1. Predictions of academic performance using ability and high school achievement measures yield multiple correlations for the four groups ranging from .23 to .47. Almost as high relationships were obtained with high school achievement alone.

2. The correlations between the discrepancy aspiration scores and academic performance provided strong support for the specific hypotheses dealing with the reality-irreality dimension. The student who held aspirations close to previous performance, who perceived his level of reasonable performance satisfaction as lying below previous performance, who believed that he would not achieve considerably beyond what he already has or his previous effort by exerting himself to the limits of his capacity, tended to be successful in grade achievement.

3. The magnitudes of the intercorrelations among the four predictors suggested that they were measuring a common factor to a high degree.

4. The combination of the four predictors in a multiple correlation indicated that higher predictability is obtained with these relatively "nonintellectual" measures than with the ability and high school achievement variables. The highest multiple correlations, however, were obtained through the combined use of both "intellectual" and "nonintellectual" measures.

5. An attempt was made to differentiate graduates from nongraduates through the summation of the four predictors into a single index. Successful discriminations were made for the freshmen and sophomore groups, but the discriminative power of the index was considerably reduced for the juniors.

6. A further attempt was made to determine whether the discriminatory power of the aspiration index was related to the academic standing of the student. This analysis revealed that the index was unsuccessful in differentiating graduates from nongraduates among below C students but was effective among the C and above C students.

In general, the results provide support for the fruitfulness of the level of aspiration method in predicting academic performance and attrition among a superior group of students. It should be emphasized that the success of the aspiration measures provides empirical confirmation of the utility of the level of aspiration approach rather than offering a practicable instrument for selection. Since the nature and setting of the particular sample employed was unique, one might not expect that such nonintellectual variables would play as meaningful a role among less homogeneous and less academically oriented students.

REFERENCES

Cronbach, L. J. Further evidence on response sets and test design. *Educ. Psychol. Measmt.*, 1950, **10**, 3–31.

Gould, R., & Kaplan, N. The relationship of "level of aspiration" to academic and personality factors. *J. soc. Psychol.*, 1940, **11**, 31–40.

Henry, E. R. Predicting success in college and university. In D. H. Fryer & E. R. Henry (Eds.), *Handbook of applied psychology*. Vol. 2. New York: Rinehart, 1950.

Holt, R. R. Level of aspiration: Ambition or defense? *J. exp. Psychol.*, 1946, **36**, 398–416.

Lewin, K., Dembo, T., Festinger, L., & Sears, Pauline S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. I. New York: Ronald Press, 1944.

Research report to the self-study committee. Unpublished manuscript. Reed College, 1953.

Rotter, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, Inc., 1954.

Schultz, D. G., & Riccuitti, H. N. Level of aspiration measures and college achievement. *J. gen. Psychol.*, 1954, **51**, 267–275.

Sears, P. S. Levels of aspiration in academically successful and unsuccessful children. *J. abnorm. soc. Psychol.*, 1940, **35**, 498–536.

Received March 22, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 2, 1959

SOME CHARACTERISTICS OF HIGH SCHOOL PUPILS FROM THREE INCOME GROUPS

JOHN K. COSTER

Department of Education, Purdue University

In a recent study, a sample of 878 high school pupils was divided into three income groups, and responses to 27 items reflecting attitudes toward high school were analyzed to ascertain variations among income groups. The attitudinal items pertained to such objects as school, teachers, school program, other pupils, and the value of education. The groups varied significantly on only eight of the 27 items. Greatest variations generally were noted on items dealing with interpersonal relationships (e.g., Do you feel that the other students like you?). There was little or no variation in responses to items which suggested an appraisal of the school or school program, such as the number of school subjects and activities offered in the school, the grading system, the administration of the school, the method of teaching, and school equipment and facilities.

After the forementioned study was completed, an analysis was made of other available data from the same sample to compare the three income groups on several items of personal information—such as sex, schooling of parents, and participation in school and community activities. The purposes of the second analysis were to identify the characteristics on which pupils in the three groups differed significantly, and to compare affective behavior (denoted by responses to attitudinal items) with factual descriptions of the *Ss*.

Several studies of the characteristics of pupils of diverse socioeconomic groups have been reported in the literature. Hollingshead (1949), for example, found that adolescents in the middle and lower-upper social classes were more likely to attend athletic events, school dances, evening plays and parties, and participate in extracurricular activities than adolescents in lower classes. Smith (1945) reported a high degree of relationship between participation in extracurricular activities and Sims Socio-Economic scale scores. In a study of junior high school pupils, Abrahamson (1952) found that middle class pupils participated in more extracurricular activities, received more scholarship awards, scored higher on social acceptance scales, held more student offices, and participated more frequently in student government, proportionately, than pupils in lower social classes.

An investigation of factors related to social acceptance of pupils living in a small community, the population of which was expanded and changed radically when a war industry was located in the area, was conducted by Morgan (1946). His findings disclosed that the social position of pupils was stabilized rapidly, and that father's income and parents' education were the two most pertinent factors bearing on social acceptance and reputation. Keisler (1954) examined the relationship between parental occupation and membership in YMCA and YWCA sponsored clubs, and reported that occupation of parents, grade point average, and scholastic aptitude were related significantly to membership in these clubs.

PROCEDURE

A questionnaire, containing 27 attitudinal items and a number of personal information items, was administered to approximately 3000 pupils in nine Indiana high schools (Coster, 1955). A sample of 878 cases was selected from the returns. The

TABLE 1
NUMBER AND PERCENTAGE OF PUPILS IN EACH INCOME GROUP

Number of items checked on House and Home scale	Number and percentage of pupils		Income group
	N	%	
0–2	219	24.9	Low
3–4	558	63.6	Middle
5–7	101	11.9	High
Totals....	878	100.0	

sample included 100 questionnaires, selected randomly, from each of six larger schools, and all usable questionnaires from three smaller schools.

The personal information items included a "House and Home" scale, designed to yield an indication of income level. This scale has been used extensively by Remmers and others in the *Purdue Opinion Panel* studies to divide pupils into income groups.[1] Remmers and Kirk (1953) and Elias (1944) have reported on the validity of the scale. The scale consisted of seven items found in the home or provided for the pupil.[2] Based on the number of items checked by each respondent, *Ss* were divided into high (five to seven items checked), middle (three or four items), and low (zero to two items) income groups. The number and percentage of pupils in each income group are shown in Table 1.

The personal information items included such categories as sex, size of family, schooling of parents, hours spent working or studying at home, grade received in school, participation in high school and out-of-school activities, and offices held in organizations. Each item in the questionnaire was followed with a list of appropriate responses, one of which was to be checked.

For each item a null hypothesis was postulated: There is no difference in the proportion of pupils checking each response category when they are divided into three income levels. Hypotheses were tested by the chi-square technique with tests based on a series of $3 \times n$ contingency tables. Adjacent response categories were combined, whenever necessary, to provide a minimum expected frequency of five in each cell. Additional combinations were made to conserve space in the tables, where reductions in the number of cells did not alter the character of the results.

RESULTS

Analyses were made of 23 personal information items. In marked contrast to the relative homogeneity of responses to the items in the attitudinal study (Coster, 1958), the pupils in the three groups varied sharply on two thirds of the personal information items. A total of 16 hypotheses were rejected, 13 at the .001 level of confidence, and one each at the .01, .02 and .05 levels.

For the report of the results, the items are grouped into four divisions: Personal Information, Success in School, Participation in School and Community Activities, and Work or Study.

[1] The *Purdue Opinion Panel* is published periodically by the Division of Educational Reference, Purdue University. Permission to use the "House and Home" scale was granted kindly by the publishers of the *Purdue Opinion Panel*. Special acknowledgement is due R. L. Horton, who suggested the items used in this scale.

[2] The items were a vacuum cleaner; an electric or gas refrigerator; a bath tub or shower with running water; two automobiles (excluding trucks); lessons in drama, art, expression, dancing, or music provided outside the school; an automatic dishwasher; and a cabin or cottage for vacations.

Personal Information

The personal information items include sex, broken homes, size of family, schooling of father and mother, and employment of mother outside the home. Most of the items in this division are associated with income level or socioeconomic status. Results of tests of significance are shown in Table 2. Three of the six hypotheses were rejected at the .001 level of confidence, one was rejected at the .02 level, and two hypotheses failed to be rejected.

Size of family and schooling of parents varied significantly with income level, with $P < .001$. Computed chi-square values greatly exceeded the value required for significance at this level. With regard to family size, high income families were more than twice as likely to have one to three children as low income families, and low income families were more than twice as likely to have four or more children as high income families. Concerning schooling of parents, variations were noted at each level of education, with greatest variations observed in the incidence of college attendance. Approximately one of four high income fathers attended college, as compared with one middle income father in eight, and one low income father in 30. Mothers varied similarly. One high income mother in four attended college, as compared with one middle income mother in seven, and one low income mother in 30.

There were variations in the percentage of boys and girls in each income group, and the attending hypothesis was rejected at the 2% level. The sample included 450 boys and 428 girls, but the difference in proportion of boys and girls was not significant. The percentage of boys exceeded

TABLE 2

RESULTS OF TESTS OF SIGNIFICANCE OF PERSONAL INFORMATION ITEMS, SHOWING PERCENTAGE OF PUPILS CHECKING RESPONSE CATEGORIES BY INCOME GROUPS

Item	Income group				
	High	Middle	Low	Total	*P*
1. Sex.					
a. Boy	48.5	54.8	43.4	51.3	
b. Girl	51.5	45.2	56.6	48.7	.02
2. With whom do pupils live?					
a. With father and mother	85.0	84.2	77.6	82.7	
b. Others	15.0	15.8	22.4	17.3	.10
3. Number of children in family.					
a. One to three children	68.3	52.7	29.6	49.5	
b. Four or more children	31.7	46.2	70.3	50.5	.001
4. Schooling of father.					
a. Grade school	18.8	36.2	56.0	39.0	
b. High school	52.5	50.8	40.8	48.5	
c. College	28.7	12.9	3.2	12.4	.001
5. Schooling of mother.					
a. Grade school	9.9	21.8	44.4	26.0	
b. High school	64.3	64.5	52.2	61.5	
c. College	25.7	13.3	3.3	12.5	.001
6. Mother employed outside of home.					
a. Yes	31.7	32.4	23.9	30.2	
b. No	68.3	67.6	76.1	69.8	.10

the percentage of girls in the middle income group, whereas the percentage of girls exceeded the percentage of boys in the high and low groups. Additional chi-square tests were made, comparing the high group with the others, the high and middle groups, the low group with the other two groups, and the low and middle groups. A significant chi-square value was obtained when the low group was compared with either the middle group, or with the other two groups. These data probably suggest a higher dropout rate among low income boys than low income girls.

A higher percentage of pupils from the low income group than from the high and middle groups lived with only one parent, a step-parent, or other relatives or persons. The differences in proportions, however, were not sufficiently large to be statistically significant. And the data in Table 2 indicate that a higher percentage of high and middle income mothers are employed outside the home, but, again, the differences were not significant.

Success in School

The "Success in School" items, listed in Table 3, included queries on plans to continue education, failure in school, and average marks. Six hypotheses were tested, and five were rejected at the .001 level. The proportion of pupils who indicated

TABLE 3

RESULTS OF TESTS OF SIGNIFICANCE OF SUCCESS IN SCHOOL ITEMS, SHOWING PERCENTAGE OF PUPILS CHECKING RESPONSE CATEGORIES, BY INCOME GROUP

Item	Income Group				
	High	Middle	Low	Total	*P*
7. Plan to graduate from high school?					
a. Yes	99.0	96.4	89.0	94.9	
b. Undecided or no	1.0	3.6	11.0	5.1	.001
8. Plan to go to college?					
a. Yes	49.5	32.1	15.5	30.0	
b. Undecided or no	50.5	67.9	84.5	70.0	.001
9. Years failed in grade school?					
a. None	94.1	86.1	77.6	84.9	
b. One or more	5.9	13.9	22.4	15.1	.001
10. Subjects failed in high school?					
a. None	93.1	87.9	82.8	87.3	
b. One or more	6.9	12.1	17.2	12.7	.05
11. Average marks received in high school.					
a. A or B average	71.3	49.4	31.2	47.4	
b. C, D or F average	28.7	50.5	68.8	52.6	.001
12. Cited on honor roll during past year.					
a. Yes	42.4	35.3	20.3	32.3	
b. No	57.6	64.7	79.7	67.3	.001
13. Field of specialization.					
a. College preparatory or academic	34.0	19.1	9.0	18.5	
b. General	13.4	17.7	13.2	16.2	
c. Vocational	52.6	63.2	77.8	65.3	.001
14. Selected life's work.					
a. Yes	42.6	34.9	32.7	35.2	
b. Not definitely or no	57.4	65.1	67.3	64.8	.30

that they plan to graduate from high school and go to college was significantly higher for the high group than for the other two groups, with $P < .001$ in both instances. There were more failures in grade school and high school subjects for the low income group than the other two groups. Differences in grade school failures were significant at the .001 level. The three groups, however, were more homogeneous with regard to failure in high school subjects, with $P < .05$.

Scholastic success, as measured by average marks in school, varied markedly among the three groups. High income pupils were more than twice as likely to report A or B grade averages than low income pupils. Similar variations were noted when pupils were asked whether their names have appeared on their school honor rolls.

Closely related to the success in school items were the questions on field of specialization and selection of life's work. Results of tests of significance for these two items are included in Table 3. Field of specialization varied significantly with income group, with $P < .001$. A larger percentage of high income pupils than middle or low income pupils were enrolled in the college preparatory or academic curriculums, whereas low income pupils were attracted in greater proportions to one of the vocational programs. Among the three groups, there was very little difference in the percentages of pupils who elected a general program. The relatively high percentage of pupils who were enrolled in a vocational program (including industrial arts in addition to vocational agriculture, homemaking, and industrial education) probably is an indication of limited curricular offerings in small schools. Practically all boys in the smaller schools were enrolled in vocational agriculture, and a large proportion of girls in the smaller schools studied homemaking for at least two or three years.

Approximately one third of the *Ss* indicated that they definitely had selected their life's work. A higher percentage of the "Yes" responses was from the high income group than from the other two groups, but the differences were not significant.

Participation in School and Community Activities

All five hypotheses pertaining to school and community activities were rejected, four at the .001 level and one at the .01 level, according to data in Table 4. One indication of participation in school activities—and social acceptance—is the extent to which youths associate with fellow pupils. A significantly higher proportion of low income pupils, than high and middle income pupils, reported that they had most of their social activities outside of school with youths who go to other schools or who are not in school.

Extremely wide variations occurred in the number of high school and out-of-school activities in which pupils of the three groups participated. Nearly ten times as many low income as high income pupils did not participate. Whereas nearly three fourths of the high income pupils reported that they participated in three or more high school activities, only one fourth of the low income group reported such extensive participation. And whereas nearly two thirds of the high income pupils indicated that they had participated in three or more out-of-school activities, less than one fifth of the low income group indicated similar participation. A divergence was also noted in the percentages of pupils who have been elected to an office. More than three fourths of the high income pupils and more than two thirds of the middle income pupils have been elected to an office in a high school or out-of-school activity, as compared with only 40% of the low income group.

Differences in the behavior of youths

TABLE 4

RESULTS OF TESTS OF SIGNIFICANCE OF PARTICIPATION IN SCHOOL AND COMMUNITY ACTIVITIES ITEMS, SHOWING PERCENTAGE OF PUPILS CHECKING RESPONSE CATEGORIES, BY INCOME GROUPS

Item	Income Group				
	High	Middle	Low	Total	*P*
15. With whom do pupils associate?					
a. With pupils in the same high school	82.2	84.1	74.0	81.4	
b. With youth in other high schools or not in school	17.8	15.9	26.0	18.6	.01
16. Participation in high school activities.					
a. None	2.0	9.5	19.2	11.1	
b. One or two activities	25.8	44.2	55.7	45.0	
c. Three or four activities	35.6	29.5	21.0	28.1	
d. Five or more activities	36.6	16.7	4.1	15.8	.001
17. Participation in out-of-school activities.					
a. None	3.0	11.3	27.9	14.5	
b. One or two activities	32.0	49.1	52.5	48.1	
c. Three or four activities	39.6	30.9	17.3	28.5	
d. Five or more activities	24.7	8.6	2.3	8.9	.001
18. Held office in an organization.					
a. Yes	78.0	69.7	40.2	63.6	
b. No	22.0	30.3	59.8	36.4	.001
19. Frequency of Church or Sunday School attendance.					
a. Three or more times per month	78.0	66.8	52.1	64.5	
b. Less than three times a month	21.0	28.7	32.8	28.8	
c. Never	1.0	4.5	15.0	6.7	.001

were noted in extent of Sunday School and Church attendance. Slightly more than three fourths of the high income pupils and over two thirds of the middle income pupils reported regular Sunday School and Church attendance, as contrasted with slightly more than one half of the low group. Fifteen per cent of the low income pupils indicated that they never attended Sunday School or Church, as compared with less than five per cent of the middle group and only one per cent of the high group.

Work and Study

The fourth division pertains to an analysis of income group variation with regard to the number of hours spent working or studying outside of school. This aspect of the study has received limited attention in other investigations of income level and social status. To get the necessary information, pupils in the sample were asked these questions:

1. About how many hours do you study at home each week?

2. About how many hours each week (during the school year) do you work at home?

3. Do you receive pay or allowance for the work you do at home?

4. About how many hours each week do you work at a job away from home during the school year from which you earn some money?

The results of tests of significance are shown in Table 5. High income pupils tended to spend more time studying at home than low and middle income pupils, and middle income pupils were more likely

TABLE 5

RESULTS OF TESTS OF SIGNIFICANCE OF WORK OR STUDY ITEMS, SHOWING PERCENTAGE OF PUPILS CHECKING EACH RESPONSE, BY INCOME GROUP

Item	Income Group				
	High	Middle	Low	Total	*P*
20. Hours pupils study at home each week.					
a. Less than five hours per week	62.4	73.4	70.4	71.5	
b. Five hours or more per week	37.6	26.6	29.6	28.5	.10
21. Hours pupils work at home each week.					
a. Less than ten hours per week	61.4	61.6	59.8	61.1	
b. Ten hours or more per week	38.6	38.3	40.3	38.9	.90
22. Receive pay or allowance for work at home					
a. Yes	42.6	50.9	42.7	47.9	
b. No	57.4	49.1	57.3	52.1	.10
23. Hours pupils work outside of home each week.					
a. Less than ten hours per week	73.3	79.0	75.9	77.6	
b. Ten hours or more per week	26.7	20.9	24.1	22.4	.50

to receive pay and allowance for the work they did at home than youths in the other two groups, but the differences were not significant. There was practically no difference among the three groups in the number of hours per week that pupils worked outside of school, either at home or away from home.

DISCUSSION

Some general comments based on a synthesis of the findings of the two studies presented in this paper and in the previous article (Coster, 1958) are here appropriate. High income pupils responded more favorably than middle or low income pupils with regard to the majority of items in both studies; however, there were marked variations in the number of items for which null hypotheses were rejected. In the present study, significant differences were obtained for the majority of the items. Responses to attitudinal items which involved an appraisal of the school and school program generally did not manifest these marked differences in school and community experiences.

There are indications that the low income pupil has tended to accept his lot in school in a state of resignation. Perhaps he is conscious of the efforts of his school to provide an appropriate educational program for him. The findings show, nevertheless, that he participates in few school activities, yet complained no more about the number of activities than high or middle income pupils. He tended to receive low grades in school, yet he was as satisfied with the grading system as high and middle income pupils whose average grades were much higher. And his opinions of the subjects offered in the school did not differ from the opinions of pupils from higher income homes, yet the low income pupil is more likely to deprecate his chances of getting the kind of job he wants after high school even though there is a greater possibility that he is enrolled in a vocational program designed to fit him for useful employment.

Responses to attitudinal items dealing

with social acceptance, in contrast, varied similarly with the school and community experience items reported in this study. It would seem, perhaps, that low income pupils recognize that there are activities in which they may participate, but that either they do not wish to participate, possibly because of variations in social values attached to these activities, or they feel that they are not welcome to join.

The findings of these two studies tend to suggest that although schools may subscribe to a position of equality of educational opportunity, equality of educational experiences falls somewhat short of the ideal. Even though a predominant theme of the contemporary philosophy of American education is to provide appropriate educational opportunities for all American youth, sociological and psychological factors tend to operate against the attainment of this goal. Then, too, criticisms to the effect that American education has been more concerned with the education of the "masses" at the expense of the "elite" are more apropos when leveled at theory than at practice. The findings from these and other studies of social status and income level suggest that American schools, consciously or unconsciously, have been prone to favor pupils from middle and upper socioeconomic strata which, generally, include a high proportion of academically competent youth.

Summary

In this study, 878 pupils were divided into three income groups, and the pupils in the groups were compared on a number of items of personal information—such as sex, schooling of parents, and participation in school and community activities. High income pupils were more likely than middle and low income pupils to participate in high school and out of school activities, hold an office in an organization, get high marks in school, be named to the school honor roll, attend Sunday School and Church regularly, successfully complete courses in school, and continue education. The number of hours spent studying or working outside of school, during the school year, however, did not differ among the groups. The number of items related to school and community experience which varied significantly with income level contrasted markedly with the relative homogeneity of responses of the three groups to attitudinal items pertaining to the school and school program, as reported in a previous study of the same *Ss* (Coster, 1958). But there seemed to be an essential agreement between differences in responses to attitudinal items dealing with social acceptance and the extent of participation in school and community activities.

REFERENCES

Abrahamson, A. Our status system and scholastic rewards. *J. educ. Sociol.*, 1952, **25,** 441–450.

Coster, J. K. Factors related to morale in secondary schools. Unpublished doctoral dissertation, Yale Univer., 1955.

Coster, J. K. Attitudes toward school of high school pupils from three income levels. *J. educ. Psychol.*, 1958, **49,** 61–66.

Elias, G. A study of certain methods of attitude measurement and related variables. Unpublished master's thesis, Purdue Univer., 1944.

Hollingshead, A. B. *Elmtown's youth.* New York: John Wiley, 1949.

Keisler, E. R. Differences among adolescent social clubs in terms of members' characteristics. *J. educ. Res.*, 1954, **48,** 297–303.

Morgan, H. G. Social relationships of children in a war-boom community. *J. educ. Res.*, 1946, **40,** 271–286.

Remmers, H. H., & Kirk, R. B. Scalability and validity of the socio-economic status items of the *Purdue Opinion Panel. J. appl. Psychol.*, 1953, **37,** 384–386.

Smith, P. A study of the selective character of American education: Participation in school activities as conditioned by socioeconomic status and other factors. *J. educ. Psychol.*, 1945, **36,** 229–246.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 2, 1959

MEASUREMENT CHARACTERISTICS OF RECALL IN RELATION TO THE PRESENTATION OF INCREASINGLY LARGE AMOUNTS OF MATERIAL[1]

WILSE B. WEBB[2] AND MARVIN SCHWARTZ[3]
U. S. Naval School of Aviation Medicine

In testing for differences in the amount of information individuals possess about a given body of information a well-known principle exists: the greater the number of items of information tested for, the greater the consistency in the measurement of individual differences. This is neatly summarized in the Spearman-Brown formula. This is *not* the point of concern of this paper. Our problem, so far as we can tell, has not received attention in the testing of individual differences. We are concerned with the relationship between the consistency of measurement of information obtained by individuals and the number of items of information presented the individuals at one presentation. In less formidable terms, would there be greater consistencies of differences between individuals who were presented, for example, 10 items of information and tested on five of these items or between individuals presented 50 items of information and tested on five of these items?

PROCEDURE

Material Presented

Six stories, averaging 741 words, edited from Bulfinch's *Mythology* (1931), were used for our information presented the *Ss*. These stories possessed the advantages of not only being generally unfamiliar to the *Ss*, but, in addition, each contained considerable detail which permitted the relatively easy development of comprehension tests (Webb & Wallon, 1956).

Tests Used

Forty-five true-false test items were employed. These items were chosen by item analyzing a considerably larger pool of items, eliminating nondiscriminative items (Webb & Wallon, 1956).

Subjects

The *Ss* in this experiment were 345 naval aviation cadets. In general they possessed a minimum of two years of college and were further screened on an intelligence type test prior to entering naval aviation.

Method of Presentation

The *Ss* (ranging in *N* from 31–69) in each group were distributed a set of stories and questions face down on their desks. They were told to read the stories and answer the questions as if they were listening to the story. They were told to read steadily forward at a comfortable rate, not going back or pausing on either the stories or the questions.

EXPERIMENTAL GROUPS

Group I

The questions relevant to each paragraph of three stories were inserted immediately behind each paragraph. The task of the *S* was merely to read a single paragraph and to reproduce the information obtained immediately afterwards.

[1] Opinions and conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department.

[2] Now at University of Florida.

[3] Now at State University of Iowa.

Group II

For this group the questions were behind each of the three stories. These *Ss* read approximately 700 words before being questioned on the information contained therein.

Group III

This group read all three stories with the questions appearing at the end of the group of three stories. Here the *Ss* read approximately 2200 words before answering the questions.

Group IV

Six stories were presented this group before the questions about the stories were presented.

In the first three groups the same three stories and 45 test items were involved. In the fourth group the addition of three different stories and sets of questions was required. The stories common to the previous three groups appeared first as did the common test items when the questions were asked.

Results

The basic data of this experiment, given in Table 1, include the means, *Ss* variance, error variance, and the resultant Hoyt homogeneity estimate of reliability for the 45 items in common between the four groups.

TABLE 1

Results of Hoyt Analysis of Variance Estimate of Reliability

Group	*N*	Mean	*S* Variance	Error Variance	Reliability
I	69	40.14	.135	.084	.378
II	127	38.20	.343	.102	.703
III	90	36.65	.454	.119	.740
IV	59	34.07	.458	.150	.672

Discussion

Our data indicate that an increase in the information presented resulted in an increase in both the between *Ss* variance on test materials and an increase in the error variance. In our experimental situation the error variance tends to increase in positively accelerated fashion whereas the *Ss* variance appears to increase in a negatively accelerated fashion. Because the *Ss* variance is considerably greater than the error variance, this results in an increasing reliability estimate through the first three conditions but a decreased estimate in the fourth condition.

It is likely that the critical factor in the increasing *Ss* variance is the increase in item difficulty and hence the discrimination between *Ss*. That there was such an increase in item difficulty can be seen from the means given in Table 1. Whether in addition the increase in difficulty resulted in a change in the factor structure and a resultant increase in the individual differences cannot be tested in this design but represents a real possibility.

The increased error variance is probably attributable to a number of factors. By extending the amount of information which must be learned, there is an increase in the probability of fluctuations in attention, variations in amounts of previous information, methods of learning, proactive and retroactive inhibition effects and failures to recall leading to guessing behavior. All of these factors could result in errors in response unrelated to the true *S* variance and should increase as the amount of information which an *S* must learn and recall increases.

Our results imply that if *Ss* are presented a limited amount of material and immediately tested for comprehension, although the error variance will be low, so too will be the *S* variance and hence the measure will have a limited reliability. As

the amount of information to be learned is increased, the between *Ss* variance will increase only up to a point but the error variance will continue to increase. Because of this, it follows that beyond a given amount of information, the classical estimates of reliability of measurement of information comprehended will decrease.

Summary

Ss were permitted to read complex stories approximately 700 words in length. In one group, questions about the stories were inserted behind each paragraph, in a second group the questions appeared behind each story, in a third group the questions followed three stories and in the fourth group the questions followed six stories. It was found that the *Ss'* variance increased in a negatively accelerated fashion with increasing amount of materials being presented prior to testing. However, the error variance also increased but in a linear or positively accelerated fashion. This relationship resulted in the reliability estimate of the scores increasing through the first three conditions and then decreasing.

REFERENCES

Bulfinch, T. *Mythology*. New York: J. M. Vent, 1931.

Webb, W. B., & Wallon, E. J. Comprehension by reading versus hearing. *J. appl. Psychol.*, 1956, **40**, 237–240.

Received August 12, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 2, 1959

PERSONALITY SYNDROMES AND ACADEMIC ACHIEVEMENT

GEORGE MIDDLETON, JR.[1] AND GEORGE M. GUTHRIE
Pennsylvania State University

Attempts to predict academic achievement have utilized three classes of variables: intelligence test scores, indices of previous achievement, and measures of personality and other nonintellective factors. To date, aptitude test scores combined with an index of high school performance in a multiple regression equation have yielded the best estimates of the grade point average. A ceiling of about .70 has been reached in these efforts, however, with most *R*s in the 50's. Attempts to improve prediction by using nonintellective factors, such as interest and personality traits, have yielded quite discouraging results. The principal sources of difficulty appear to lie in the heterogeneity of the criterion and the nonsummative and nonlinear properties of many promising predictors.

It is worthwhile to examine the most frequently used criteria, the grade point average, and pass-fail. While they provide a convenient continuum or dichotomy, they are reached by subjects by many different routes. Ghiselli (1956) has discussed the parallel problem that arises from the use of one over-all numerical rating as a criterion of performance in an industrial setting. Averaging grades frequently entails averaging uncorrelated values. One behavior pattern, such as conformity, may be rewarded in one course and another of independence in a different course. The availability and the social significance of the averages on transcripts have encouraged persons doing prediction studies to overlook the heterogeneous origins of the values which contribute to these criterion scores.

[1] Now at McNeese State College, Lake Charles, Louisiana.

Even if we are predicting achievement in a homogeneous subject matter area, or for a single course, we must take into account the several sets of motivational structures which further or impede achievement. Worrying, for example, may facilitate one student's work and cause another to fail. Some light on this can come from the unsystematized observations of persons with similar achievement but different abilities and of those with similar aptitude scores and very divergent performance. It is clear that there are many patterns of combining aptitudes, methods of studying, and motivation to achieve a similar grade point average. This has led some clinicians to resort to nonmathematical combining of predictors in an effort to reach a more satisfying global assessment of the individual's likelihood of success. For all that this method may provide more convincing after-the-fact explanations of success or failure, Meehl's summary of statistical and clinical prediction (1954) shows that combining indices "in our heads" does not yield as good results as even relatively simple statistical methods. And neither do very well!

In a break with traditional designs, Frederiksen and Melville (1954) were able to demonstrate that correlations of a predictor variable may be increased in some cases by including a moderator variable, a variable which bears no relationship to the criterion. Specifically, they demonstrated that taking out compulsive *S*s will improve prediction of the other *S*s, even though compulsiveness is not correlated with either the predictor or the criterion. Saunders (1956) has presented a mathematical basis of moderated regression and further evi-

dence of its operation. This work suggests that personality factors may be associated with achievement but not necessarily in a linear or curvilinear relationship. Nor is the relationship that of the usual suppressor variable. Rather, there may be at least one subgroup, and probably more than one, characterized by certain personality characteristics in which the relationship between intelligence and achievement is different from that among those not in the subgroup.

A method of obtaining groups of similar individuals has been demonstrated by Stephenson (1953, pp. 190–218). By inverse factor analysis he separated high and low achievers on a series of items regarding study habits. This approach is discussed by Stern, Stein and Bloom who provided the rationale for this study. Since the same criterion score can be achieved in a variety of ways, the multiple regression model may be inappropriate. This study utilizes their suggestion that, "A possible technique for dealing with non-homogeneous populations ... is provided by transposed factor analysis" (Stern, Stein, & Bloom, 1956, p. 235). In this way, it may be possible to identify homogeneous subgroups with different personality syndromes and then to seek the factors leading to different levels of academic achievement.

Method

An attempt has been made to delineate the personality syndromes among high achieving and low achieving students. The *Ss* were male students in a single curriculum. From a pool of about 50 business management students who had attained at least junior standing, 14 with averages of 2.50 or higher where A is 4.00, and 14 with averages below 2.00 were selected. These groups had done equally well in high school but the high group had significantly higher scores on a college aptitude test.

The *Ss* were given a 300-item personality questionnaire, Table A,[5] made up of items drawn from Murray (1938, pp. 142–242). These items had been designed to measure 18 of the needs in Murray's system. Each item was answered true or false. Examples of the items are:

I enjoy psychological novels more than other kinds of literature.
I became very attached to my friends.
I am intolerant of people who bore me.

Adjusted phi coefficients were computed between each pair of *Ss*. The 14 × 14 matrices for high and low achievers were factor analyzed, and the rotated factor loadings were correlated with the scores on each of Murray's 18 provisional scales to permit interpretation of the factors.

Results

The results are shown in Table 1 and in Tables B through K.[5] Five factors were extracted from the matrix for high achievers. Ten of 14 *Ss* showed loadings of at least .40 on one or more factors. These factors were interpreted by referring to their correlations with the 18 provisional scales. The factors are described in terms of the needs in Murray's system which correlated at least .40 with the factor.

Factor H-I. This factor correlates positively with nurturance and dominance and negatively with abasement, succorance and narcissism. Achievement for these *Ss* appears to mean power and approval.

Factor H-II. Autonomy, aggression, counteraction, and achievement correlate positively, while abasement and affiliation are negatively correlated. Achievement for

[5] Tables A–K have been deposited with the American Documentation Institute. Order Document No. 5854, remitting $2.00 for 35 mm microfilm or $3.75 for 6 by 8 in. photocopies. Order from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Wash. 25, D.C. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

CORRELATIONS BETWEEN ROTATED FACTORS AND PROVISIONAL SCALES FOR HIGH AND LOW ACHIEVERS

Need[a]	High Achievers					Low Achievers			
	I	II	III	IV	V	I	II	III	IV
Intra	.25	−.13	.47	.09	−.46	.60	.17	−.61	.06
n-Agg	−.17	.51	−.03	−.25	.43	.36	−.22	−.35	.32
n-Sen	−.04	−.41	.63	−.26	−.23	.83	−.10	−.79	.07
n-Exh	−.17	−.45	.78	−.05	−.35	.47	−.21	−.62	.55
Extra	.10	.27	−.31	.49	−.05	−.34	.45	.81	−.05
n-Aff	.15	−.45	.53	.36	−.78	.15	.41	.00	.43
n-rej	.23	.32	−.15	−.09	.15	.20	−.43	.38	.20
N	−.52	.00	.22	−.33	.31	.64	−.38	−.68	−.39
n-Suc	−.52	−.31	.80	.00	−.42	.82	−.68	−.55	−.14
n-Nur	.52	−.40	.22	.52	−.78	.06	.52	−.17	.40
n-Inf	−.25	−.23	.27	.50	−.52	.75	−.42	−.34	−.76
n-Ctn	.34	.46	−.29	.32	.05	−.96	.38	.31	.29
n-Und	.23	−.23	−.11	.15	−.49	−.10	.52	.06	.52
n-Dom	.42	−.17	.21	.41	−.37	.10	.10	−.10	.54
n-Auto	−.13	.54	.08	−.29	.20	.24	−.08	−.71	−.06
n-Def	.09	−.36	.45	.55	−.76	.58	−.18	−.36	−.06
n-Aba	−.89	−.59	.64	.00	−.11	.63	−.55	−.45	−.58
n-Ach	.15	.43	.04	.07	−.09	.16	−.08	.34	−.20

[a] The needs are identified as in Murray (1938).

this group seems to be an expression of resentment and independence.

Factor H-III. The needs of succorance, exhibition, abasement, sentience, and affiliation all correlated positively with this factor. In contrast to the preceding factor, this factor presents strong dependence.

Factor H-IV. Deference, nurturance, infavoidance or avoidance of failure, extraception, and dominance correlated positively with Factor H-IV. These *Ss* appear to be pursuing goals of social prestige and influence. Achievement may be an avenue whereby they can be thought well of.

Factor H-V. Aggression shows a low positive correlation, while high negative correlations appear with nurturance, affiliation, deference, infavoidance, and understanding. Achievement appears to be related to a hostile aggressive denial of tender socialized feelings.

These five factors must be regarded as more illustrative than confirmed. However, the results do suggest that achievement of high grades may be motivated by drives for power, resentment, dependence, social acceptance, and aggression. These findings are supported by anectodotal accounts of counselors.

Four factors were extracted from the matrix of low achievers. Eleven *Ss* showed correlations of at least .40 with one or more factors.

Factor L-I. The following needs showed high correlations with this factor: sentience, succorance, infavoidance, narcissism, and abasement, while counteraction or a need to overcome defeat was highly negatively correlated. These *Ss* appear to be preoccupied with pleasures.

Factor L-II. Nurturance, understanding, extraception and affiliation are positively correlated with this need, while succorance, abasement, rejection, and infavoidance are

denied. These persons appear to be insistently extroverted in their relationships.

Factor L-III. Extraception is the only positively correlated need, while sentience, autonomy, narcissism, exhibitionism, and succorance are negatively related to this factor. This group seems intent on disavowing social shortcomings.

Factor L-IV. Needs of exhibition, dominance, understanding, and affiliation are positively related to this factor, while the *S* seeks to avoid blame and threats to his self-esteem. These *S*s appear to be preoccupied with power and acceptance. The four low achieving factors reflect trends toward pleasure seeking, extroversion, denial of normal shortcomings, and power. Again, these make sense, after the fact, in students who are doing poorly in school.

The important point about these results is not confirmation of specific syndromes, but rather the demonstration that these independent factors do exist. Their confirmation would point to the next step of seeking improved predictors for each group or the use of these factors to qualify predictions. Beyond this, remedial work with failing students may be improved by a clarification of the systems of attitudes that appear frequently among them.

Summary and Conclusions

Fourteen high achieving and 14 low achieving students answered a 300-item personality questionnaire. The matrices of phi coefficients between pairs of persons were factor analysed. Five factors of persons were found in the matrix of high students and four factors were found in the other matrix. Although the factors need much further confirmation, they are highly suggestive. The method employed will permit identification of homogeneous groups of students for whom academic achievement may play a similar role. Improved prediction and modified counseling techniques should be made possible by clearer recognition of different syndromes of personality which are associated with different levels of achievement.

REFERENCES

Frederiksen, N., & Melville, S. D. Differential predictability in the use of test scores. *Educ. psychol. Measmt.*, 1954, **14**, 647–656.

Ghiselli, E. E. Dimensional problems of criteria. *J. appl. Psychol.*, 1956, **40**, 1–4.

Meehl, P. E. *Clinical versus statistical prediction*. Minneapolis: Univer. Minnesota Press, 1954.

Murray, H. A. *Explorations in personality*. Oxford Univer. Press, 1938.

Saunders, D. R. Moderator variables in prediction. *Educ. psychol. Measmt.*, 1956, **16**, 209–222.

Stephenson, W. *The study of behavior*. Univer. Chicago Press, 1953.

Stern, G. G., Stein, M. I., & Bloom, B. S. *Methods in personality assessment*. Glencoe: Free Press, 1956.

Received August 28, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 2, 1959

PREDICTING PRINCIPALS' RATINGS OF TEACHER PERFORMANCE FROM PERSONALITY DATA

ROBERT F. PECK
University of Texas

Teacher personality is known to influence teaching effectiveness (Barr, 1955; Hunt, 1956; Ryans, 1952; Symonds, 1955). In an effort to analyze this relationship further, a pilot study was undertaken using projective personality data obtained from 49 teachers, from five school systems, who had been nominated by their principals as high, average, or low on five aspects of teacher behavior (Stiles, 1957). As an exploratory measure of the relationship between personality and teaching performance, an attempt was made to predict the principals' ratings from independent analysis of the personality data on the teachers.

To venture to predict the evaluation that will be given to different teachers, in different school systems, by different principals whose individual attitudes and expectations are not known, might well be regarded as a venture in foolhardiness. To make such an attempt does require the assumption that there are some fairly universal standards for judging teaching effectiveness, despite the many situational factors that weigh somewhat differently with each school and each community (Ryans, 1957). Another assumption is also implied: that school principals, whatever their differences of opinion, share enough values in common in appraising teachers, and are stable enough in their judgments, that their evaluations can be predicted with reasonable accuracy if the individual personalities of the teachers are known. Finally, of course, it is a considerable assumption to make that the personalities of teachers can validly be assessed from projective data. Such a study was undertaken, nonetheless, as an exploratory test of both concepts and instruments which might prove useful in a five-year research program now getting under way.[1]

PROCEDURE

The Teacher Effectiveness Scale

In order to measure several aspects of the teaching role which presumably are somewhat independent of one another, a five-part criterion instrument was constructed. School principals were asked to nominate a teacher who was high, one who was average, and one who was low on each of these five scales:

I. Organizing and communicating information and skills.

II. Creating a healthy relationship with pupils.

III. Creating good relations with other teachers.

IV. Building good relations in the community.

V. Supervisor's personal evaluation ("Who would you pick to take with you if you moved to a new school?").

A descriptive paragraph was given with each scale, to provide some concrete illustrations of the intended meaning of the scale.

[1] This study was part of a research program of the Laboratory of Human Behavior, Department of Educational Psychology, University of Texas—a program financed by funds from the Hogg Foundation for Mental Hygiene. Relationships among college experiences, individual mental health, and teaching effectiveness are currently under study at the Mental Health Demonstration Center, University of Texas, under a grant from the National Institute of Mental Health.

The Projective Instruments

The teachers nominated by the principals were asked to complete two forms and return them directly to the research office: a four-page Biographical Information form; and the Peck Sentence Completion, Form 2-B, a 90-item instrument designed for use with educational personnel and with college students of education.

Steps in Data Collection

Eighteen elementary schools were contacted by letter and then by personal call, to invite participation in the study.[2] Ten principals agreed to take part, and provided Teacher Effectiveness nominations. Every precaution was taken to insure that the principals' ratings were known only to the principal and to the research director and his statistical assistant. A systematic effort was made to obtain a second, independent assessor in each school, such as a curriculum supervisor; but this turned out to be not feasible.

Each participating principal gave a set of the biographical and sentence completion forms to each of the teachers he or she nominated on any of the Teacher Effectiveness scales, with the invitation to contribute this information to a study of teachers' attitudes and interests. Some indication that participation was generally voluntary may be found in the fact that of the 10 schools participating, complete data from both principal and teachers were obtained from only five schools. Forty-nine teachers from these five schools thus constitute the final research sample.

The five cities from which teachers provided data ranged in size from 5,000 through 15,000 and 30,000 to 500,000. This gave a wide range and diversity in size of school system and in cultural milieu. Geographically, the cities spanned the state from end to end, further diversifying the cultural contexts of the schools. One was a rural seat, one a heavily industrialized town, and one a highly metropolitan city.

The 49 teachers ranged in age from 22 to 70, with a median age of 42: 10 were 30 years of age; 12 were in their 30's, 15 were in the 40's; 11 were in the 50's; and one was 70; all but two were women. In experience, they ranged from one to 50 years of teaching, with a median experience of 14 years. The great majority were married. Some had taught continuously since the beginning; some had entered teaching after children reached school age; and some had taught, withdrawn to marry, then resumed teaching 6, 8, or even 20 years later. Considering the high attrition rate in sample size, it is as much good luck as planning that the final sample was quite heterogeneous.

It should be noted that there was no significant relationship in any of these five subsamples between age or experience and the effectiveness ratings.

Analysis of the Projective Data

The biographical and sentence completion forms were coded for school and individual, then separated into five groups according to whichever of the five Teacher Effectiveness scales each teacher had been nominated on. This gave subsamples as follows: Scale I: 13 *Ss*; II: 11; III: 11; IV: 13; V: 10.

A "blind," qualitative analysis was made of each combined pair of Biographical and Sentence Completion protocols, and brief notes were recorded on each teacher's personality characteristics. (See Peck & Thompson [1954] for analogous procedure in another set of studies of personality

[2] Credit is due Ralph Duke, Carson McGuire, A. C. Murphy, John Newell and George White for indispensable help in this phase of the study, particularly. Credit due the participating principals and teachers can only be expressed in general, so that identities are protected.

and job performance.) Following this, a comparison was made of the people within the subsample for the scale under investigation, and an effort was made to classify each teacher as high, average, or low on that scale. There was some confusion frequently as to whether this rating should represent the judge's own appraisal of the person, against the scale as the judge visualized it, or whether it should represent an effort to match the principal's probable rating, regardless of the judge's personal opinion. A compromise was generally attempted, leaning toward an effort to match the principal's supposed rating. In some cases, however, particularly on Scale I, the judge found there was a limit beyond which he felt impelled to record his own evaluation, although he suspected that the principal's rating wound differ. The groups were analyzed and rated in this order: I, III, IV, II, V, I.

A product-moment correlation was separately computed for each group. For this purpose, the "principal's" rating was treated as a score (e.g., "High" = 1, "Low" = 3), and the ratings from the projective data were similarly treated. *N*s were small, ranging in size from 10 to 13 *S*s in the various scale-groups. No reliability measure was attempted, since no second judge appropriately trained in the use of these particular instruments was free to participate in the study, at the time. At most, therefore, positive findings could only serve as encouragement to more extensive research, with more specific delineation of the analytic procedures.

Results

The correlation of principals' ratings with projective-based ratings, for each scale, was as follows:

I. Organizing and communicating information and skills. The original correlation between the projective analyst's rating and the five principals' ratings was —.42,[8] a nonsignificant relationship.

II. Creating a healthy relationship with pupils. The correlation here was .19, insignificant.

III. Creating good relations with other teachers. The projective analyst's ratings correlated .82 with the principals' ratings, a relationship significant beyond the .01 level.

IV. Building good relations in the community. The correlation was .89, significant yond the .01 level.

V. Supervisor's personal evaluation. The correlation of the projective-based rating with the principals' was .84, significant beyond the .01 level.

Discussion

Knowing nothing about the principals, the schools or the specific communities involved, the analyst was able to "predict" the ratings of five different principals on three of the Teacher Effectiveness scales, with an almost surprisingly high degree of accuracy. Yet, curiously, the same analyst, when appraising for classroom teaching ability and for effect on pupil's mental health, disagreed with the principals (and about equally with all five). This discrepancy calls for examination.

The first explanation, of course, might be that the analyst is a poor judge of classroom teaching, and of mental health. Undoubtedly this is true to some extent, although there is a certain amount of evi-

[8] Upon rerating this group of teachers at the end of the study, the analyst knew that his original judgment was almost opposite in direction to the principals' so he tried deliberately to force himself to rate as the principals would. On several cases, however, he concluded it would do violence to the nature of the data, and of the teaching process, to do other than record his own assessment of the teacher. The final correlation with the principals' ratings was .42, a nonsignificant relationship.

dence to the contrary (Peck & Thompson, 1954; Peck & Parsons, 1956). Apparently the analyst can understand and match school principals' ratings with considerable accuracy, when it comes to judging teachers' effectiveness in relating to other teachers, to the community, and to the principal.

A second interpretation might be that these particular kinds of data permit prediction of the latter three kinds of behavior, but not of the first two. This seems quite unlikely, however, in view of the analytic process that was followed. An over-all picture of each teacher's personality pattern was built up from the data by the analyst. Then, from this picture were inferred the various kinds of behavior to be rated. There is no immediately visible reason why it should prove possible to infer complex relations with an (unknown) community accurately, yet impossible to infer relationships with children, in the much better known setting of the classroom.

Another possible source of error, accounting for the lack of agreement on Scales I and II, has to do with possible incompleteness or inaccuracy of the principals' knowledge of what goes on inside the classrooms. Where the principals have frequent opportunities to observe teachers at first hand—as in the situations Scales III, IV, and V represent—they may achieve an accurate picture of the teachers' relative efficiencies in these settings. When the principals have relatively little opportunity to see the teachers' natural behavior, as is actually the case with what happens inside the classroom in many schools, then the principals may not really have an adequate basis for judging how effective particular teachers actually are, in classroom action.

Indeed, some indications in the data suggest that the principals may unconsciously be influenced by factors quite apart from the teachers' classroom efficiency. Several of those rated high on Scales I and II by the principals, give evidence in the projective data of a mild but pervasive neurasthenia, combined with a quietly firm self-restraint amounting almost to self-effacement. These women did not seem to the analyst to have either the emotional aliveness or the sheer energy to be able to attend alertly and respond effectively to either the intellectual or emotional states and changes of individual children. These teachers have an organized way of keeping things quiet, however; and perhaps this is what marks them in some principals' eyes, rather understandably, as good teachers: they keep quiet, keep out of the way, demand little attention, and meanwhile keep the children from "getting in the principal's hair."

This may not be either a true or a fair interpretation in the present instance (although some of the accurate matching of the principals' ratings on the other scales proceeded from somewhat similar reasoning). It is important to raise for detailed inquiry, however, the question of whether efficient custodians are not sometimes mistaken for effective teachers by principals too harried to note the distinction. And, of course, there may be principals who actively prefer and seek to have quietly submissive teachers, a possibility which would raise some question about the mental health and emotional climate of those schools.

When one considers the room for error at any one of the several stages of the study, let alone all of them together, perhaps the wonder is that any degree of accuracy of prediction could be achieved. As is always true of the half-science, half-art of human assessment, it requires a complex series of "educated guesses"—requires it, because the number, nature and interaction of all the factors that affect such a thing as teaching effectiveness are far too intricate for adequate representation by a few "pure" measures of single variables or by

even the best mathematical models available at present.

Summary and Conclusions

Forty-nine experienced elementary school teachers from five geographically separated cities in Texas were rated by their principals on a Teacher Effectiveness instrument. A given teacher was nominated as being high, average, or low on one of five scales: I. Organizing and communicating information and skills; II. Creating a healthy relationship with pupils; III. Creating good relations with other teachers; IV. Building good relations in the community; V. Supervisor's personal evaluation.

A projective personality analysis was made of each teacher, on the basis of a biographical form and a sentence completion form. From this, an effort was made to match the criterion ratings. Correlations of the projective ratings with principals' ratings were: Scale I, .42; II, .19; III, .82; IV, .89; V, .84. Highly significant prediction was achieved on the last three scales. Possible reasons for the disagreements on intraclassroom effectiveness were discussed. The positive results suggest that there may be some fairly universal standards for judging teacher effectiveness; that various principals' judgments are stable and valid enough, at least where the principals have a chance for first-hand observation, to be predictable; and that personality can be validly appraised from projective data (Alexander, 1950). The results appear favorable to further experimentation with the kind of data and projective analysis employed in this study.

REFERENCES

Alexander, T. Prediction of teacher-pupil interaction with a projective test. *J. clin. Psychol.*, 1950, **6**, 273–276.

Barr, A. S. Measurement and prediction of teacher efficiency. *Rev. educ. Res.*, 1955, **25**, 261–269.

Hunt, J. T. School personnel and mental health. *Rev. educ. Res.*, 1956, **26**, 502–521.

Peck, R. F., & Parsons, J. Personality and work output. *Personnel Psychol.*, 1956, **9**, 49–79.

Peck, R. F., & Thompson, J. The use of individual assessments in a management development program. *J. personnel Admin. industr. Relat.*, 1954, **1**, 79–98.

Ryans, D. G. A study of criterion data (A factor analysis of teacher behaviors in elementary school). *Educ. psychol. Measmt.*, 1952, **12**, 333–344.

Ryans, D. G. Notes on the criterion problem in research, with special reference to the study of teacher characteristics. *J. genet. Psychol.*, 1957, **91**, 33–61.

Stiles, L. J. *The teacher's role in American society*. New York: Harper, 1957.

Symonds, P. M. Characteristics of the effective teacher based on pupil evaluations. *J. exp. Educ.*, 1955, **23**, 289–310.

Received August 28, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 2, 1959

THE RELATIONSHIP OF HOME AND SCHOOL EXPERIENCES TO SCORES ON ACHIEVEMENT TESTS[1]

JOHN W. FRENCH
Educational Testing Service

In connection with college entrance, an important part is played by standard validity studies designed to find out how well aptitude or achievement tests are able to predict academic grades. Through such studies much can be learned about the necessary qualifications for college work, but it is also of interest to know what background situations and what experiences affect the test scores. Such information may have some implications for college preparation, but, more importantly, it can improve an understanding of the ways in which students acquire the knowledges and skills that are being measured.

The tests used in this experiment were developed to provide information on a student's preparation for college but with increased emphasis on ability to apply knowledge of principles to solutions of problems and decreased emphasis on knowledge of specific details. There are six tests in all: Science Glossary which concerns scientific language and concepts; Science Abilities which concerns the solution of scientific problems; Social Studies Abilities calling for understandings in the field of history; Social Studies Essay using the essay approach to the same thing; Humanities References covering familiarity with ideas in the fields of literature, music, and art; and Humanities Abilities covering an understanding of literature, music, and art.

The tests were administered to high school seniors at 26 private and 15 public schools. To permit comparisons between the experimental tests and a standard test, only those students who had taken the College Board's Scholastic Aptitude Test were included in the study. The students also filled out a questionnaire on family background, courses taken, courses liked, hobbies, reading habits, and their own opinion of how tests should rate their ability or achievement in various fields. The schools supplied grades and special ratings by teachers. Complete data were obtained for 1,275 boys and 725 girls.

Preliminary Screening Analysis

Because the variables were quite numerous, it was desirable to eliminate from further study those that had no important relationships. For each of the two or more categories defined by each variable, the number of *Ss* making low, middle, and high test scores were tabulated by the IBM 101 machine. To avoid capitalization on chance all tabulations were made separately for two random halves of the data, referred to later as Sample 1 and Sample 2.

Inspection of the tables gave no evidence of curvilinearity and revealed no important discrepancies between the two samples. The tables were then used for the elimination from further study of questionnaire, rating, or experience variables having low relationships with any of the test scores. Unexpectedly, father's occupation and education showed little or no relationship to the test scores, and were, therefore, eliminated. Hobbies in music, art, mechanics, and sports were also left out, as were the number of books read in mechanics and in sports and the number of courses taken in art, music, religion, home economics, business, and vocations. Variations in the num-

[1] The study was supported by the College Entrance Examination Board in connection with the tryout of a series of experimental tests entitled Tests of Developed Ability.

ber of courses taken in English would undoubtedly affect some of the test scores, but this was omitted as a variable, because almost all *Ss* took four years of high school English. A liking expressed for some courses was found to be important, but the variables pertaining to a liking for art, music, religion, home economics, business, and vocational courses were eliminated.

Having taken physics and having taken chemistry are two variables which did show high relationships with the science tests but were omitted from further study, since it was thought that the over-all number of science courses taken would represent physics and chemistry well enough. It is notable that neither Biology nor General Science courses was substantially related

TABLE 1

CORRELATIONS OF EXPERIMENTAL TESTS AND SAT SCORES WITH RATING AND EXPERIENCE VARIABLES FOR BOYS

	Variables	Sci. Gloss.	Sci. Abil.	S.S. Abil.	S.S. Essay	Hum. Ref.	Hum. Abil.	SAT-V	SAT-M
Tests	SAT-V	.56	.69	.79	.49	.75	.69		.71
	SAT-M	.61	.77	.62	.38	.53	.55	.71	
School Ratings	High School Decile	.46	.48	.46	.35	.43	.39	.48	.51
	Science Grades	.58	.52	.40	.24	.25	.29	.39	.45
	Social Stud. Grades	.25	.32	.54	.35	.42	.33	.46	.33
	Humanities Grades	.43	.38	.44	.30	.45	.41	.48	.41
	Science Rating	.60	.58	.39	.25	.30	.34	.42	.46
	Social Stud. Rating	.20	.31	.54	.42	.49	.42	.45	.30
	English Rating	.33	.43	.48	.40	.56	.52	.66	.45
Self Ratings	Science Estimate	.65	.59	.30	.15	.17	.28	.33	.44
	Math. Estimate	.52	.58	.35	.14	.23	.32	.35	.71
	Social Stud. Estimate	.00	.07	.50	.29	.35	.20	.33	.06
	English Estimate	.23	.29	.42	.32	.48	.44	.55	.27
	Art Estimate	.08	.08	.06	.01	.15	.17	.11	−.04
	Music Estimate	.06	.07	.02	.03	.16	.23	.07	−.02
	Language Estimate	.18	.19	.27	.26	.42	.37	.40	.32
Experience Variables	Science Hobby	.39	.39	.12	−.01	.06	.13	.12	.20
	Social Hobby	.17	.19	.14	.15	.24	.13	.18	.23
	Writing Hobby	.11	.18	.39	.27	.43	.38	.41	.23
	Science Read	.47	.38	.18	.04	.13	.12	.24	.23
	Mechanics Read	.19	.14	−.06	−.10	−.09	−.08	−.07	.07
	History Read	.02	.04	.28	.16	.24	.11	.16	.13
	Government Read	−.10	−.05	.24	.13	.24	.10	.19	.02
	Travel Read	.00	−.02	.10	−.03	.05	.04	.03	−.15
	Fiction Read	.13	.15	.32	.18	.32	.30	.35	.20
	Music or Art Read	.01	.05	.10	.03	.29	.14	.15	.03
	Science Courses	.23	.07	−.13	−.10	−.14	−.16	−.10	.03
	Math. Courses	.38	.39	.01	−.01	.00	.05	.05	.40
	Social Stud. Courses	−.31	−.27	−.15	−.07	−.22	−.18	−.26	−.28
	Language Courses	.11	.17	.34	.16	.45	.48	.46	.35
	Science Liked	.39	.33	.02	−.02	−.02	.03	.06	.08
	Math. Liked	.32	.37	.04	.06	.03	.10	.02	.42
	Social Stud. Liked	−.20	−.17	.22	.18	.17	.04	.08	−.14
	English Liked	−.07	−.05	.18	.26	.28	.22	.25	−.08
	Language Liked	−.05	−.02	.10	.09	.33	.25	.20	.16

Note.—The figures are averaged tetrachoric correlations for Samples 1 and 2.

to performance on the science tests. A count of the test items from the various fields of science does not explain this finding.

In the field of social studies no course was found to be substantially related to the tests. Variability in taking American history would presumably have shown a relationship to the social studies tests, but almost all students had taken the course.

Results for Selected Variables

This paper will be mainly concerned with what we can call experience variables: hobbies, reading, courses taken, and courses liked. Courses liked are considered to be ex-

TABLE 2

CORRELATIONS OF EXPERIMENTAL TESTS AND SAT SCORES WITH RATING AND EXPERIENCE VARIABLES FOR GIRLS

	Variables	Sci. Gloss.	Sci. Abil.	S.S. Abil.	S.S. Essay	Hum. Ref.	Hum. Abil.	SAT-V	SAT-M
Tests	SAT-V	.48	.68	.81	.49	.74	.73		.57
	SAT-M	.43	.71	.51	.37	.41	.44	.57	
School Ratings	High School Decile	.32	.49	.54	.29	.44	.41	.59	.55
	Science Grades	.36	.34	.34	.25	.30	.24	.35	.44
	Social Stud. Grades	.24	.43	.45	.22	.36	.40	.52	.46
	Humanities Grades	.32	.48	.50	.27	.45	.48	.65	.39
	Science Rating	.51	.56	.55	.38	.36	.36	.55	.55
	Social Stud. Rating	.29	.40	.60	.30	.45	.41	.61	.44
	English Rating	.37	.48	.62	.32	.57	.55	.68	.48
Self Ratings	Science Estimate	.43	.41	.15	.20	.12	.12	.31	.35
	Math. Estimate	.35	.54	.25	.20	.14	.19	.25	.72
	Social Stud. Estimate	.03	.16	.43	.29	.32	.21	.41	.04
	English Estimate	.24	.32	.48	.23	.47	.47	.66	.28
	Art Estimate	.06	−.06	.01	−.04	.15	.16	.08	−.23
	Music Estimate	−.11	−.12	−.08	−.06	.15	−.01	−.02	−.03
	Language Estimate	.15	.16	.37	.24	.34	.17	.38	.26
Experience Variables	Science Hobby	.22	.24	.15	.10	.07	.09	.12	.01
	Social Hobby	.06	.17	.18	.11	.09	.13	.19	.25
	Writing Hobby	.15	.18	.31	.19	.35	.27	.41	.18
	Science Read	.25	.12	−.16	−.10	−.09	−.04	−.07	.02
	Mechanics Read	.16	.03	.09	.04	.02	.15	.08	−.02
	History Read	.14	.05	.18	.02	.25	.22	.20	.05
	Government Read	.01	.07	.19	.02	.23	.21	.25	.11
	Travel Read	.11	.01	.17	.10	.27	.11	.21	.08
	Fiction Read	.08	.14	.25	.15	.29	.24	.35	.18
	Music or Art Read	.12	.12	.20	.07	.30	.25	.25	.08
	Science Courses	.44	.35	.07	−.08	.00	.07	.07	.10
	Math. Courses	.07	.15	−.08	−.19	.06	.05	.04	.36
	Social Stud. Courses	−.11	−.09	.03	−.09	.03	−.08	.00	−.18
	Language Courses	.11	.07	.22	.17	.20	.16	.24	.21
	Science Liked	.33	.15	−.08	−.09	−.09	−.05	−.08	−.06
	Math. Liked	.25	.31	.08	.02	−.14	−.08	−.04	.52
	Social Stud. Liked	−.01	−.04	.30	.23	.21	.14	.14	−.12
	English Liked	.10	.14	.25	.09	.25	.36	.37	.06
	Language Liked	.12	−.01	.21	.07	.19	.06	.17	.04

Note.—The figures are averaged tetrachoric correlations for Samples 1 and 2.

perience variables as they indicate very important qualitative aspects within a student's school experience. As mentioned earlier, education and occupation of the father, which are also experience variables, were omitted, because they were found to bear only a slight relation to test scores as compared to the other variables just mentioned. Table 1 for Boys and Table 2 for Girls give tetrachoric correlations between selected variables and scores on the experimental tests and Scholastic Aptitude Test (verbal and mathematical). Suffice it to say here that correlations of the experimental test scores with school grades, teachers' ratings on ability, and self ratings, were all suitably high for corresponding tests and subject-matter fields.

Analysis. Since the tests can be evaluated in one way by the extent to which they measure skills and knowledge picked up from appropriate experiences, we can consider one kind of validation for the tests to be the multiple correlation of the relevant experience variables with scores on each test. Here we have the reverse of the situation existing in many prediction studies. Instead of a single criterion and multiple predictors, we have a single predictor and multiple criteria. The criteria may be weighted either by *judgment* according to what kinds of experiences each test *should* reflect, or they may be weighted so as to be *maximally predictable*. Both of these methods of weighting were tried.

TABLE 3

MULTIPLE CORRELATIONS OF EXPERIMENTAL TEST SCORES WITH CRITERIA CONSISTING OF EXPERIENCE VARIABLES WEIGHTED ACCORDING TO JUDGMENT

Tests	Multiple Correlations[a]			
	Boys		Girls	
	Sample 1	*Sample 2*	*Sample 1*	*Sample 2*
Science Glossary	.556	.504	.477	.430
Science Abilities	.401	.438	.360	.274
Social Studies Abilities	.197	.248	.399	.144
Social Studies Essay	.106	.194	.130	.008
Humanities References	.516	.483	.489	.375
Humanities Abilities	.391	.395	.510	.315

[a] Formula after Kelley taken from Dunlap and Kurts (1932),

$$r_{xc}\left(\sum_{1}^{A} W_y X_y\right) = \frac{r_c}{\sqrt{\frac{1 - r_i}{A} + r_i}}$$

where r_c is the mean validity, r_i the mean intercorrelation among predictors, and A the number of predictors.

Judgment Criterion. For want of any more insightful judgment, this criterion will consist of an equal weighting of all experience variables in the restricted field of the test. These variables are listed as follows:

Science	Social Studies
Science Hobby	History Read
Science Read	Government Read
Science Courses	Social Studies Courses
Science Liked	Social Studies Liked

Humanities
Writing Hobby
Fiction Read
Music or Art Read
English Liked

Table 3 presents the multiple correlations based on equal weightings of the appropriate four variables. The higher multiple correlations may be interpreted as instances where the test measures what it was *judged* that it *should* measure. By this criterion the Science Glossary test looks best; the social studies tests look poorest.

Maximally Predictable Criterion. Separately for each of the three fields, each sex, and each sample, variables were selected to constitute a most predictable criterion. Table 4 lists the tests selected for Sample 1, gives the beta weights, and gives both the multiple correlations for Sample 1 and the multiple cross-validation, that is, the figure obtained when the Sample 1 weights were applied to Sample 2. Table 5

TABLE 4

MULTIPLE CORRELATIONS OF EXPERIMENTAL TEST SCORES WITH EXPERIENCE CRITERIA WEIGHTED FOR MAXIMUM PREDICTABILITY, WEIGHTING BASED ON SAMPLE 1

Test	Experience Variables	Beta-Weights and Multiple Correlations					
		Boys			*Girls*		
		β	Validation (Sample 1)	Cross-Val. (Sample 2)	β	Validation (Sample 1)	Cross-Val. (Sample 2)
Science Glossary	Science Hobby	.039			−.060		
	Social Hobby	.177			—		
	Science Read	.365			.320		
	Mechanics Read	—			.125		
	Math. Courses	.275	.654	.578	—	.656	.465
	Math. Liked	−.043			.180		
	English Liked	—			.233		
	Science Courses	—			.389		
	Science Liked	.196			—		
Science Abilities	Science Hobby	.105			.225		
	Social Hobby	.141			—		
	Science Read	.263			.024		
	Mechanics Read	—			.057		
	Math. Courses	.184	.530	.539	—	.618	.478
	Math. Liked	.091			.335		
	English Liked	—			.272		
	Science Courses	—			.311		
	Science Liked	.047			—		
Social Studies Abilities	Writing Hobby	.283			.233		
	Fiction Read	.109			.074		
	History Read	.058			—		
	Government Read	—	.547	.524	.132	.549	.375
	Music or Art Read	—			.153		
	English Liked	−.073			.112		
	Language Courses	.318			—		
	Soc. Stud. Liked	.145			.150		
Social Studies Essay	Writing Hobby	.164			.140		
	Fiction Read	.158			.130		
	History Read	−.029			—		
	Government Read	—	.408	.301	.026	.307	.266
	Music or Art Read	—			−.002		
	English Liked	.298			.007		
	Language Courses	−.029			—		
	Soc. Stud. Liked	−.080			.160		
Humanities References	Writing Hobby	.313			.077		
	Fiction Read	.091			.043		
	History Read	—			−.013		
	Government Read	.064	.655	.581	.216	.526	.390
	Music or Art Read	—			.233		
	English Liked	.054			.253		
	Language Courses	.384			—		
	Language Liked	.050			—		

TABLE 4 (*Continued*)

Test	Experience Variables	Beta-Weights and Multiple Correlations					
		Boys			*Girls*		
		β	Validation (Sample 1)	Cross-Val. (Sample 2)	β	Validation (Sample 1)	Cross-Val. (Sample 2)
Humanities Abilities	Writing Hobby	.361	.635	.587	.054	.535	.411
	Fiction Read	.083			.117		
	History Read	—			−.094		
	Government Read	−.120			.128		
	Music or Art Read	—			.268		
	English Liked	.094			.310		
	Language Courses	.524			—		
	Language Liked	−.176			—		

does the same thing with the samples reversed.

The validation figures in Tables 4 and 5 tend, of course, to be higher than the cross-validation figures, since they take full advantage of chance fluctuation in the correlation coefficients. The cross-validation figures represent a best estimate of the true correlation of the tests with a maximally predictable combination of experience variables. It will be seen that these tend to be higher than the correlations of the tests with the judgment criteria given in Table 3. There can be two reasons for this: (*a*) the judgment of what the tests *should* measure was imperfect, or (*b*) the tests measure something different from what they *should* measure.

It is particularly true for the social studies tests that the judgment validities are much lower than the validities based on empirically computed weights. This implies that the social studies tests are related to some of the experience variables but not very closely to the ones to which they might be expected to relate or *should* relate. Inspection of the beta weights in Tables 4 and 5 shows that number of historical books read, number of books on government read, and number of social studies courses taken do not have the expected importance in predicting the scores on the social studies tests. Liking social studies courses is an important contributor as expected, but the other good ones include having a writing hobby, number of language courses taken, and number of books of fiction read. This seems to mean that the social studies tests are too highly related to the purely verbal or communication abilities rather than to the more specialized abilities or understandings needed in the social studies field. However, the correlations in Tables 1 and 2 suggest that, if verbal variance as defined by the verbal section of the Scholastic Aptitude Test were partialled out of the social studies test scores, beta weights for the number of books read about history and government would become more prominent.

Discussion and Conclusions

Although we are studying experience variables with the intention of judging the effect of experience on test scores, we cannot assume that high test scores for a certain group are necessarily *caused* by the special experiences of that group. For example, Table 1 shows a positive correlation between number of language courses taken and all of the experimental tests. What evidence is there that the language experience directly affected the test scores?

TABLE 5

MULTIPLE CORRELATIONS OF EXPERIMENTAL TEST SCORES WITH EXPERIENCE CRITERIA WEIGHTED FOR MAXIMUM PREDICTABILITY, WEIGHTING BASED ON SAMPLE 2

Test	Experience Variables	Beta-Weights and Multiple Correlations					
		Boys			*Girls*		
		β	Validation (Sample 2)	Cross-Val. (Sample 1)	β	Validation (Sample 2)	Cross-Val. (Sample 1)
Science Glossary	Science Hobby	.076	.607	.599	.215	.656	.596
	Writing Hobby	.274			.219		
	Science Read	.249			—		
	Math. Courses	.202			—		
	Math. Liked	.071			.158		
	Language Courses	—			.338		
	Science Courses	—			.454		
	Science Liked	.233			.134		
Science Abilities	Science Hobby	.162	.643	.584	.170	.549	.590
	Writing Hobby	.288			.283		
	Science Read	.144			—		
	Math. Courses	.200			—		
	Math. Liked	.200			.306		
	Language Courses	—			.242		
	Science Courses	—			.401		
	Science Liked	.175			−.110		
Social Studies Abilities	Writing Hobby	.295	.534	.536	.065	.427	.439
	Social Hobby	—			.166		
	Fiction Read	.099			.155		
	History Read	.187			−.022		
	Government Read	−.043			—		
	Language Courses	.166			.147		
	Soc. Stud. Liked	.198			.237		
Social Studies Essay	Writing Hobby	.222	.385	.331	.098	.309	.312
	Social Hobby	—			.081		
	Fiction Read	−.028			.040		
	History Read	.127			−.010		
	Government Read	−.060			—		
	Language Courses	.145			.089		
	Soc. Stud. Liked	.225			.212		
Humanities References	Writing Hobby	.245	.555	.593	.231	.457	.452
	Fiction Read	.106			.056		
	History Read	—			.122		
	Music or Art Read	.188			.117		
	English Liked	.053			.013		
	Language Courses	.264			.210		
	Language Liked	.012			—		
Humanities Abilities	Writing Hobby	.228	.534	.562	−.010	.450	.425
	Fiction Read	.189			−.041		
	History Read	—			.268		
	Music or Art Read	−.011			.022		
	English Liked	−.015			.300		
	Language Courses	.299			.218		
	Language Liked	.084			—		

Let us assume that the special experiences being considered in this study have relatively little effect on the Scholastic Aptitude Test. This is a reasonable assumption, because these tests are known to be relatively noncoachable. (Dear, 1958; Dyer, 1953; French, 1955) This assumption makes it possible to get some idea of the kind of student who undergoes each experience. Reference again to Table 1 shows that the number of language courses taken has a very substantial correlation with Scholastic Aptitude Test scores. This suggests that the language experience may have had no effect on the experimental test scores, since the high correlation between these tests and the number of language courses taken can be accounted for by the fact that the students electing five years of languages are well above average in verbal and mathematical aptitude.

On the other hand, the number of science courses taken correlates well with the science experimental scores, but does not correlate appreciably with either section of the Scholastic Aptitude Test. In this case, since the experimental scores are not accounted for by the general scholastic aptitude of students electing science, it seems probable that the science courses or some other experiences associated with students who take four years of science are responsible for the high test scores. Partial correlations were used to demonstrate this effect in more detail, but the figures obtained are not essential to this article.

The results of this study indicate that science experience is well represented in the tests. In connection with the variable, number of science courses taken, the preliminary screening study suggests that the effect on the test scores is present for physics and chemistry but not for general science and biology. Since general science is almost universally taken by high school students, little discrimination by this variable could be expected. The low relationship with taking biology is less easy to understand, since biology was well represented among the test items.

The humanities tests, as might be expected, have high relationships with writing and reading activities. However, their lack of relationship with music and art hobbies confirms what the nature of the tests seems to indicate: namely, that scholarly appreciation and knowledge of music and art were being measured rather than talent or skill.

The social studies tests are so highly verbal that they also measure writing and general reading activities rather than experience specific to the social studies. A real challenge in test construction will be to produce a test of understanding in the field of social studies which is not too highly verbal. The low relationship with courses other than American history is a finding that is consistent with the emphasis to be found in the test items.

The results of this study can be taken as a demonstration that, when a test is to be evaluated in terms of its relationship to multiple experiences (or criteria), useful information on what the test actually measures can be had by computing the multiple correlation with experience variables and cross-validating if possible. However, the only way to evaluate how well the test measures what it *should* measure requires use of weights that have been decided upon in advance.

REFERENCES

DEAR, R. E. The effects of a program of intensive coaching on SAT scores. Princeton, N. J.: Educational Testing Service, 1958. (Unpublished report.)

DUNLAP, J. W., & KURTZ, A. K. *Handbook of statistical nomographs, tables, and formulas*. New York: World Book, 1932.

DYER, H. S. Does coaching help? *Coll. Bd. Rev.*, 1953, No. 19, 331.

FRENCH, J. W. An answer to test coaching. *Coll. Bd. Rev.*, 1955, No. 27, 5–7.

Received September 19, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 2, 1959

THE EFFECT OF AN INTRODUCTORY PSYCHOLOGY COURSE ON SELF-INSIGHT[1]

FRANK COSTIN
University of Illinois

Can a beginning course in psychology increase self-understanding? While instructors frequently assert that it can, or at least should, little research has been published showing that it actually does (Birney & McKeachie, 1955). The purpose of this paper is to report the results of a study designed to help answer the question.

METHOD

Four introductory psychology classes were investigated. All of these pursued the same one-semester course, which was part of a two-year general education program in the Division of General Studies at the University of Illinois. (While the Division is an integral part of the College of Liberal Arts and Sciences, its courses are open to and attract students from a wide variety of colleges within the University.) A major objective of the course was to help students understand their own behavior.

Following a one-week orientation to the nature of psychology, these topics were studied: perception (two weeks), motivation (three weeks), learning (three weeks), intelligence (three weeks), and personality (five weeks).

Classes attended lectures two hours a week, with each half of the class also meeting as a separate discussion section for an additional two hours a week. A basic text, supplementary references, and films amplified the scope of the lectures and discussions. The same instructor taught all four classes over a period of four consecutive semesters.

A 37-item "self-insight" scale, developed by Gross (1948a), was administered to each class at the beginning and end of the course. This particular instrument was selected because some studies concerned with the effects of teaching methods had indicated it had good possibilities for measuring a rather specific aspect of self-understanding (Gross, 1948b; Ruja, 1954). The scale was also given at the beginning and end of the semester to a class in verbal communication, another course in the Division of General Studies. These 97 students were similar to the psychology students in age and college representation. They had never taken a college course in psychology and were not currently enrolled in one. The reason for testing them was to see what effect general college experience, independent of a psychology course, might have on the kind of self-insight measured in this study.

The following excerpts from Gross' detailed description of his instrument explain in part its rationale: "Self-insight is the acceptance and admission of both the presence and absence of personality traits within oneself when this acceptance ... clashes with one's feelings of self-esteem" (Gross, 1948a, p. 223). "The most severe test of self-insight will be found in the ability of the individual to accept as true those truths which are implicitly or explicitly denied by social usage and to accept as false those falsehoods which are implicitly or explicitly affirmed by social usage". (Gross, 1948a, p. 222).

The following sample statements illustrate more specifically the nature of the scale: "I have criticized other people for saying things which I might very well have said myself." "Occasionally I have sexual thoughts which I would not like to reveal to other people." "I can as easily

[1] A briefer version of this paper was presented on September 1, 1958, at the convention of the APA in Washington, D. C.

laugh at myself as at other people." "I have no feeling of hostility toward anyone." According to Gross, an individual will reveal his degree of self-insight by the manner in which he responds to statements like these. For example, strong agreement with the item, "I have criticized other people for saying things which I might very well have said myself," shows a high degree of self-insight. Strong disagreement expresses low self-insight. On the other hand, strong agreement with the item, "I have no feeling of hostility toward anyone," reveals low self-insight. Strong disagreement indicates high self-insight.

Students were asked to respond to each of the 37 statements in the scale by choosing one of five options: strongly agree, agree, uncertain, mildly disagree, and strongly disagree. In the scoring procedure originally developed by Gross, a 2 or 1 was assigned to responses revealing relatively high self-insight, and a −2 or −1 to responses showing relatively low self-insight. "Uncertain" responses were scored 0. Since the present study was planned so as to use mean scores in estimating changes in self-insight, a modification of Gross' scoring system was made in order to simplify computations. This change consisted of eliminating the negative sign from −1 and −2, and changing 0, 1, and 2 to 3, 4, and 5, respectively. Thus, for example, strong agreement with the statement, "I have no feeling of hostility toward anyone," received 1; mild agreement, 2; uncertainty, 3; mild disagreement, 4; and strong disagreement, 5. Accordingly, the higher a student's total score on the scale, the greater was his self-insight.

TABLE 1

SELF-INSIGHT OF STUDENTS BEFORE AND AFTER A COURSE IN INTRODUCTORY PSYCHOLOGY AS COMPARED TO A CONTROL GROUP TAKING A COURSE IN VERBAL COMMUNICATION

	Students in Introductory Psychology	Students in Communication Course
Before Course:		
Mean	117.87	117.30
SD	16.32	13.05
After Course:		
Mean	121.63	117.11
SD	16.32	14.71
Difference in Means	3.76	−.19
t	4.37*	.17
N	179	97

Note.—High scores indicate greater self-insight.
* $P < .01$.

RESULTS

As Table 1 shows, psychology students made a significant increase in self-insight score, the mean change being 3.76 scale points. The verbal communication class, however, revealed no significant change in its mean score. A comparison of the data for the two groups shows that the initial mean self-insight scores of the psychology and the verbal communication students were approximately the same at the beginning of their respective courses. Only the psychology students, however, changed significantly. It would seem, then, that the psychology classes changed as the result of taking a course in that subject, and not simply because of general college experiences.

To discover what relationship might exist between knowledge acquired in the psychology course and increase in self-insight, students achieving in the upper half of their class were compared with those in the lower half. Achievement was measured by four objective examinations given during the semester, and a comprehensive final examination at the end of the course. Results of this analysis are described in Table 2. The self-insight scores of the two groups were approximately the same at the beginning of the course, the actual difference of 1.59 having a *t* value of .65. Upper half students increased their self-insight sig-

TABLE 2

SELF-INSIGHT OF STUDENTS ACHIEVING IN UPPER AND LOWER HALVES OF INTRODUCTORY PSYCHOLOGY

	Before Course		After Course		Difference	
	Mean Score	*SD*	Mean Score	*SD*	in Means	*t*
Upper half (N = 90)	118.66	16.69	124.01	17.08	5.35	4.42*
Lower half (N = 89)	117.07	15.89	119.21	15.14	2.14	1.75

Note.—The higher the score, the greater the self-insight.
* $p < .01$.

TABLE 3

SELF-INSIGHT OF MEN AND WOMEN ACHIEVING IN UPPER AND LOWER HALVES OF INTRODUCTORY PSYCHOLOGY

	Before Course		After Course		Difference	
	Mean Score	*SD*	Mean Score	*SD*	In Means	*t*
Upper half						
Men (N = 57)	119.37	16.34	124.70	14.80	5.33	3.63[a]
Women (N = 33)	117.42	17.22	122.82	20.37	5.40	2.49[b]
Lower half						
Men (N = 60)	117.88	15.53	121.25	15.14	3.37	2.25[c]
Women (N = 29)	115.38	16.52	115.00	14.24	.38	.20

Note.—The higher the score, the greater the self-insight.
[a] $p < .01$.
[b] $.02 > p > .01$.
[c] $.05 > p > .02$.

nificantly by 5.35 scale points. The lower half's increase of 2.14 was not a significant gain.

An analysis was also made of sex differences in self-insight change. Men and women students achieving in the upper half of their class were compared, as were men and women in the lower half. Table 3 shows these results. Upper-half men and women both increased their scores, and to the same extent. Lower-half women did not change. Lower-half men, however, increased their scores significantly, and the amount of change (3.37) did not differ significantly from that shown by upper-half men (5.33). (This difference of 1.96 had a t value of .48.) It should also be noted that the initial scores of both male groups revealed no significant difference, the t value being .51.

DISCUSSION

This study has shown that a significant increase in self-insight was made by students who completed a one-semester introductory psychology course. Since verbal communication students did not change, it is reasonable to conclude that the psychology course was the effective agent in producing change, and not simply general university experiences. Gain in self-insight apparently had a positive relationship to the acquisition of course content, as meas-

ured by classroom examinations, because the total group of students achieving in the upper half of their class, on the basis of examination scores, gained significantly in self-insight, while the total group of students achieving in the lower half did not change.

In addition to the information which was actually measured by course examinations, however, other kinds of learnings probably contributed to increased self-insight, at least in the case of one group of students—men whose scholastic achievement in the course was in the lower half of their class. This conclusion is inferred from the fact that while both men and women in the upper half of their class made significant gains in self-insight, men in the lower half also gained. This finding does not necessarily mean, of course, that the increased self-insight revealed by the lower-half males was independent of any information which they acquired from the course. More likely, it may mean that some of the knowledge which influenced them to change, although learned through taking psychology, was simply not measured by the examinations to any appreciable extent. Or, to state this in a slightly different manner, the kind of information which the psychology instructor's tests measured may not have been exclusively essential for gaining the kind of self-insight measured by Gross' scale. It is also quite possible that retention of only a relatively small amount of the course content, or acquisition of a rather restricted range of information, was sufficient to result in the attainment of self-insight.

The fact, then, that men whose scholastic achievement was in the lower half of their class did increase their self-insight makes it clear that at least for them a kind of course achievement took place that was not apparent simply by examining their subject matter test scores. Other factors, not measured in this study, must also have been influential in increasing their self-insight. This interpretation, however, must be considered again in the light of still another finding: women achieving in the lower half of their class did not change significantly. No data were collected to explain why this occurred. Some plausible speculations can be invoked, such as differences in ability and motivation between lower-half men and women. (Motivation is a more likely clue.) Further investigations should be carried out to see whether this particular sex difference was unique for this study, or whether similar sex differences would continue to be obtained. For the purpose of the present investigation, the chief value of the finding that lower-half women did not gain in self-insight is that it gives more weight to the previously stated inference: information acquired in the course, as measured by classroom examinations, played an important and specific part in influencing self-insight increase.

As stated at the beginning of this paper, a major objective of the psychology course was to help students understand themselves better. The results of the investigation reported here indicate that this objective was achieved. The actual amount of self-insight increase, however, was rather small. Or, to state this more accurately, there was much less change than the investigator had anticipated. Perhaps it was unrealistic to have expected a greater change from so brief an encounter with formal psychology. Then, too, half of the course dealt with topics which were not as likely to increase the kind of self-understanding measured as were those of motivation and personality. Had this study been carried out in a beginning course containing more material on personality and motivation, greater degrees of change might well have occurred.

How appropriate was the use of the term "self-insight" to describe the changes measured in this investigation? This particular

label was used because it was the name Gross had given to his scale. One might reasonably question whether the kinds of responses evaluated with this instrument really represented "self-insight." The concept certainly means quite different things to different psychologists; even those who would agree on its meaning might argue that its use in this study was inappropriate. (The reader will probably already have noticed the similarity between Gross' scale and the Lie Scale of the Minnesota Multiphasic Personality Inventory.) The present investigator is inclined to think that it would be more accurate to employ a specific term like "self-acceptance" or "decrease in defensiveness" to describe the kinds of changes which were measured.

Granting the soundness of the foregoing objections to the rubric "self-insight," the fact remains that an introductory psychology course *was* demonstrated to be somewhat effective in increasing a readily measurable kind of self-understanding. Since self-understanding, of one kind or another, is a goal which many instructors, and practically all students of psychology subscribe to, it would be highly desirable if further studies of this kind were made. Such research could profitably be concerned not only with an introductory course similar to that described in the present report, but with courses organized quite differently. Other aspects of self-understanding, calling for different kinds of measuring instruments, should also be investigated. (In the case of advanced courses, a more sophisticated kind of device than that used in the present study would probably be necessary.) As more information of this sort is accumulated, teachers of psychology would be in a better position to evaluate the extent to which their course goals of self-understanding are being achieved.

SUMMARY

Gross' self-insight scale was administered to 179 undergraduates before and after an introductory course in psychology, which was part of a two-year general education program. The group as a whole showed a small but significant increase in its self-insight score. Both men and women whose scholastic achievement in the course was in the upper half of their class increased their self-insight scores significantly, and to the same extent. In the lower half, men also increased self-insight significantly, but women did not change. It was concluded that changes in self-insight were positively related to information acquired from the course as measured by objective examinations, although course learnings not evaluated by the examinations also probably played a part in effecting change. As a control measure, Gross' scale was given to 97 students before and after a verbal communication course, also a part of the general education program. No significant change in self-insight score occurred.

REFERENCES

BIRNEY, R., & MCKEACHIE, W. The teaching of psychology: A survey of research since 1942. *Psychol. Bull.*, 1955, **52**, 51-68.

GROSS, L. The construction and partial standardization of a scale for measuring self-insight. *J. soc. Psychol.*, 1948, **28**, 219-236. (a)

GROSS, L. An experimental study of the validity of the non-directive method of teaching. *J. Psychol.*, 1948, **26**, 243-248. (b)

RUJA, H. Nondirective teaching and self-insight: A statistical addendum. *J. gen. Psychol.*, 1954, **51**, 331-332.

Received October 8, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 2, 1959

RETENTION IN ARITHMETIC AMONG CHILDREN OF LOW, AVERAGE, AND HIGH INTELLIGENCE[1] AT 117 MONTHS OF AGE[1]

HERBERT J. KLAUSMEIER AND JOHN F. FELDHUSEN[2]
University of Wisconsin

Considerable research in learning efficiency and retention has been completed during the last 50 years. McGeoch and Irion (1952, p. 376) cautiously summarized their review of research thus: "It seems, then, that the slow learner gains no advantage in retention from his slowness, and that the fast learner is at no disadvantage because of his superior speed." The studies reviewed by McGeoch and subsequent research generally include these features: the learning task content is nonsense syllables, unrelated words or numerals, pictures or color designs; the slow and fast learners are identified according to scores made on an initial learning task that is common to all the subjects; a variety of procedures are followed in the presentation of the acquisition task, in setting the interval between acquisition and retention, and in securing and treating the retention scores in relation to the acquisition scores.

The researchers in the present study, different from most previous researchers, were concerned with fast and slow learners as usually defined in school situations; the learning tasks were limited to those intimately related to arithmetic learning and instruction in the elementary school; the learning task was graded to the achievement level of each child; and two conditions of learning and two intervals of retention were employed. The primary purpose of this investigation was to test the hypothesis: retention of arithmetic learning is the same among children of low, average, and high intelligence at a mean age of 117 months when the original task is graded to the learner's achievement level.

[1] The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

[2] The project was under the direction of Klausmeier with Feldhusen and John Check serving as research assistants.

METHOD

Subjects. The sample consisted of 20 boys and 20 girls of low intelligence (WISC IQs 56–81), 20 boys and 20 girls of average intelligence (WISC IQs 90–110), and 20 boys and 20 girls of high intelligence (WISC IQs 120–146). The low IQ children were drawn from two special classrooms in Madison and ten special classrooms in Milwaukee. Excluded from the study were low IQ children exhibiting a second handicap to a marked degree or definite organic symptoms of retardation such as mongolism. The average and high IQ children were drawn from 11 regular classrooms in Madison. All children had birthdates between September 15, 1947 and December 15, 1948. The mean age of the children at the midpoint of the study, February 15, 1958, was 117 months.

Procedure. A counting and an addition task were developed and administered to each child individually in a room of the school building. Retention measures were secured exactly five minutes and approximately six weeks after the first acquisition measure for each child.

The procedure for the counting task is now presented. Selecting a task at each child's achievement level was accomplished by means of a survey test which involved a procedure similar to the vocabulary test of the Stanford Binet. Each child was asked

to count without assistance by 1's, 2's, 3's, 4's . . . 22's in ten steps, i.e., 3's to 30, 13's to 130, until a level was found where he made two consecutive errors, or after encouragement quit or stated he could not go farther. Ten numbers or items at this level, where he could not count correctly without assistance, constituted his new task to be learned. He was then taught the ten numbers for a period of 19 minutes. If he learned the ten numbers, e.g., 9, 18, 27 . . . 90 in less than 19 minutes, he was given overlearning activities throughout the remaining time. At the end of 19 minutes, he was asked to count aloud the ten new numbers he had learned; no assistance was given. The total correct was his acquisition score. For every child, regardless of level of task, the score could vary from 0–10. Following the acquisition test, a 5-minute reading period ensued for the child. Then, the first or 5-minute retention score was secured by having the child again count the same ten numbers aloud without assistance. The 6-week retention test was identical to the acquisition and the 5-minute retention test. The 5-minute and the 6-week retention scores in counting could also vary from 0–10 for each child.

The addition task was started for each child approximately 12 weeks after the counting task. The survey, teaching, acquisition, reading period, 5-minute retention, and 6-week retention procedures for addition were the same as for counting except for one variable. As soon as the child learned the ten addition items, teaching was stopped, the acquisition test given, and the 5-minute reading period started. Thus, in the counting task, both the time to learn (19 minutes) and the number of items to be learned (ten) were held constant for all children; this, in turn, permitted unequal amounts of time for overlearning. In the addition task, the number of items (ten) taught to and performed correctly once by the child was also held constant, but time to learn the ten items was allowed to vary up to a maximum of 17 minutes in order to prevent overlearning. This purposeful variation of the learning arrangement was done to allow testing of the hypothesis under conditions of no overlearning but unequal amount of time to learn, and of overlearning but equal time to learn.

Four additional features of the overall procedures are now clarified.

The final placement of items according to difficulty levels in the survey tests of counting and addition was done after extensive review of general research on grade placement of arithmetic learnings, of curriculum guides in arithmetic generally, of the specific guides and textbooks in the cooperating schools, and tryout with a small number of children. The counting task levels actually found through the survey test ranged from counting by 1's to 3's for the low IQ children, with pennies in some instances; for the average IQ, from 3's to 16's; and for the high IQ, from 3's to 23's. The median levels were counting by 2's, 7's, and 12's for the three groups respectively. The addition task levels for the low IQ children ranged from adding digits to adding 2-place numbers, 2 addends, with carrying; 39 of the 40 average IQ children added without assistance items involving 3 addends, 3-place numbers and 38 had as their new task level adding yards, feet, and inches; 33 of the high IQ group also had the same linear measure addition with only 3 moving into the level involving common fractions and mixed numbers.

While the child was learning to count, any error he made was called to his attention and corrected. To each of his initial correct responses, the researcher usually followed immediately with "Right," "Correct," or "Good," and with the same words to a series of correct responses once learned. In the addition task every incorrect response was called to attention and cor-

rected; every correct response also was called to the child's attention with words as above and as soon as a correct response was made, the next of the 10 items was started (for further details, see Feldhusen, 1958). Thus, all ten of the possible ten correct responses in each task were reinforced.

Since the interval between the first presentation of the counting task and the addition task was approximately 12 weeks, it was felt that little intertask effect would occur, i.e., the counting task would not affect the addition task. Had such an effect occurred, it was assumed that the results of the analysis of the addition task would not be affected in such a way as to alter the pattern of means and differences since all of the children were taught both tasks with about the same time interval between tasks.

Six members of a larger research group, including the two authors, administered the survey-teaching-acquisition-retention sequence. To assure uniformity, a small number of children were brought into the learning laboratory as instructional subjects for the research team. Weekly meetings were also held to clarify procedures during the six months the data were gathered.

Results

The means and standard deviations of the acquisition, 5-minute retention, and 6-week retention scores in the counting and the addition tasks are presented in Table 1. The mean acquisition scores of 7.55, 9.52, and 9.65 in counting and of 6.28, 5.70, and 7.38 in addition suggest that, though the *task level* may have been quite accurately identified for each child both in counting and in addition, counting a series of ten items and adding a series of ten items at the identified task level was not equally difficult, according to IQ levels. The high IQ group shows the highest mean acquisition in both tasks; however, the low IQ group is lowest in counting (7.55), while the average IQ group is lowest in addition (5.70). It was previously noted that 38 of the average and 33 of the high IQ group had the same task level in addition—the adding of yards, feet, and inches. These 71 individuals thus had ten identical addition items to learn, but the high group learned more items than did the average.

Partly because the mean acquisition scores were not the same for the three IQ groups, analysis of covariance (Edwards, 1950) was used in testing the hypothesis. Also, analysis of covariance was used because of the high correlations between acquisition and 5-minute retention scores, as

TABLE 1

Means and Standard Deviations of Acquisition, 5-Minute Retention, and 6-Week Retention Scores in Counting and Addition

IQ Group	Counting						Addition					
	Acq.		5-Min. Ret.		6-Wk. Ret.		Acq.		5-Min. Ret.		6-Wk. Ret.	
	M	σ	*M*	σ	*M*	σ	*M*	σ	*M*	σ	*M*	σ
40 Low IQ	7.55	2.53	7.35	2.90	5.28	3.11	6.28	2.78	6.18	2.92	4.80	2.83
40 Av. IQ	9.52	1.34	9.50	1.14	7.10	2.90	5.70	2.36	6.00	2.57	3.15	1.59
40 High IQ	9.65	.94	9.65	.73	7.98	2.77	7.38	1.90	7.88	1.81	5.23	3.25

TABLE 2

CORRELATIONS BETWEEN ACQUISITION AND RETENTION SCORES IN COUNTING AND ADDITION

IQ Level	*N*	Counting		Adding	
		Acquisition and 5-Min. Ret.	Acquisition and 6-Wk. Ret.	Acquisition and 5-Min. Ret.	Acquisition and 6-Wk. Ret.
Low	40	.76	.39	.89	.63
Average	40	.66	.23	.87	.27
High	40	.63	.29	.89	.67
Total	120	.79	.42	.89	.56

TABLE 3

ANALYSIS OF COVARIANCE FOR ACQUISITION, 5-MINUTE RETENTION, AND 6-WEEK RETENTION MEASURES IN COUNTING AND ADDITION WITH ACQUISITION THE COVARIATE

Analysis Area	Source of Variation	*df*	Mean Square	*F*[a]	b_{yx}
Acquisition & 5-Min. Ret.: Counting	Adjusted means	2	4.46	2.67	
	Within groups	116	1.67		.768
	Total	118			
Acquisition & 6-Wk. Ret.: Counting	Adjusted means	2	21.22	2.64	
	Within groups	116	8.05		.516
	Total	118			
Acquisition & 5-Min. Ret.: Addition	Adjusted means	2	4.66	3.24	
	Within groups	116	1.44		.916
	Total	118			
Acquisition & 6-Wk. Ret.: Addition	Adjusted means	2	18.77	3.60	
	Within groups	116	5.22		.598
	Total	118			

[a] For significance at the .05 level, *F* must equal or exceed 3.08; for significance at the .01 level, 4.80. The .01 level of significance is required for acceptance or rejection of the hypothesis.

shown in Table 2. These relatively high correlations suggest indirectly that the procedures used in surveying, teaching, and measuring acquisition and 5-minute retention were quite accurate.

The results of the four analyses of covariance for the acquisition and two retention measures in counting and in addition are shown in Table 3. With 2 and 116 degrees of freedom as shown, an *F* of approximately 3.08 is required for significance at the .05 level and of approximately 4.80

TABLE 4

ADJUSTED MEANS FOR RETENTION IN COUNTING AND ADDITION WITH FIRST ACQUISITION SCORES AS THE COVARIATE

IQ Group	Counting		Addition	
	5-Minute	6-Week	5-Minute	6-Week
Low	8.40	5.35	6.34	4.90
Average	9.03	6.78	6.69	3.60
High	9.08	7.60	7.03	4.67

for the .01 level. No F obtained is sufficiently large to be significant at the .01 level. The .01 level was required for acceptance or rejection of the hypothesis, but differences at the .05 level are reported.

The Fs of 3.24 and 3.60 are significant beyond the .05 but not at the .01 level. Table 4 shows that for addition the adjusted mean scores for 5-minute retention and 6-week retention by IQ groups were as follows: low IQ, 6.34, 4.90; average IQ, 6.69, 3.60; high IQ, 7.03, 4.67. Table 4 also gives the adjusted retention means for counting by IQ group and for 5-minute and 6-week retention as follows: low, 8.40 and 5.35; average, 9.03 and 6.78; and high, 9.08 and 7.60.

Since no F was significant at the .01 level, the researchers conclude that retention of arithmetic learning is the same among children of low, average, and high intelligence at a mean age of 117 months when the original task is graded to the learner's achievement level.

Summary

The purpose of this study was to test the hypothesis that retention of arithmetic learning is the same among children of low, average, and high intelligence when the original task is graded to the learner's achievement level. The two graded arithmetic tasks were ten items in counting and ten items in addition at a difficulty level found appropriate for each child. The conditions of learning were with time and number of items held constant in the counting task; but in the addition task time was permitted to vary to eliminate overlearning. In both tasks, the correct response to each item was reinforced during initial learning. The interval between acquisition and 5-minute retention was under complete control of the researchers, while the interval between 5-minute and 6-week retention was under no control. The interval between the counting and the addition task was 12 weeks. Using analysis of covariance, the researchers found no difference among means of the three IQ groups significant at the .01 level and conclude that retention is the same for the three groups as hypothesized.

REFERENCES

Edwards, A. F. *Experimental design in psychological research.* New York: Rinehart, 1950.

Feldhusen, J. F. A study of efficiency of learning and retention in arithmetic among children of low, average, and high intelligence. Unpublished doctoral thesis, Univer. Wisconsin, 1958.

McGeoch, J. A., & Irion, A. I. *The psychology of human learning.* New York: Longmans, 1952.

Received October 27, 1958.

MICROFORM

All journals published by the American Psychological Association are being made available on MICROFILM or MICROCARD.

Psychological Review
American Psychologist
Psychological Bulletin
Psychological Abstracts
Contemporary Psychology
Psychological Monographs
Journal of Applied Psychology
Journal of Consulting Psychology
Journal of Educational Psychology
Journal of Experimental Psychology
Journal of Abnormal and Social Psychology
Journal of Comparative and Physiological Psychology

Available only in volume units; no single issues.

For MICROFILM, order from:

UNIVERSITY MICROFILMS, INC.
313 North First Street
Ann Arbor
Michigan

For MICROCARD, order from:

J. S. CANNER & COMPANY, INC.
Microcard Division
618 Parker Street
Roxbury 20, Massachusetts

for the .01 level. No F obtained is sufficiently large to be significant at the .01 level. The .01 level was required for acceptance or rejection of the hypothesis, but differences at the .05 level are reported.

The Fs of 3.24 and 3.60 are significant beyond the .05 but not at the .01 level. Table 4 shows that for addition the adjusted mean scores for 5-minute retention and 6-week retention by IQ groups were as follows: low IQ, 6.34, 4.90; average IQ, 6.69, 3.60; high IQ, 7.03, 4.67. Table 4 also gives the adjusted retention means for counting by IQ group and for 5-minute and 6-week retention as follows: low, 8.40 and 5.35; average, 9.03 and 6.78; and high, 9.08 and 7.60.

Since no F was significant at the .01 level, the researchers conclude that retention of arithmetic learning is the same among children of low, average, and high intelligence at a mean age of 117 months when the original task is graded to the learner's achievement level.

Summary

The purpose of this study was to test the hypothesis that retention of arithmetic learning is the same among children of low, average, and high intelligence when the original task is graded to the learner's achievement level. The two graded arithmetic tasks were ten items in counting and ten items in addition at a difficulty level found appropriate for each child. The conditions of learning were with time and number of items held constant in the counting task; but in the addition task time was permitted to vary to eliminate overlearning. In both tasks, the correct response to each item was reinforced during initial learning. The interval between acquisition and 5-minute retention was under complete control of the researchers, while the interval between 5-minute and 6-week retention was under no control. The interval between the counting and the addition task was 12 weeks. Using analysis of covariance, the researchers found no difference among means of the three IQ groups significant at the .01 level and conclude that retention is the same for the three groups as hypothesized.

REFERENCES

Edwards, A. F. *Experimental design in psychological research.* New York: Rinehart, 1950.

Feldhusen, J. F. A study of efficiency of learning and retention in arithmetic among children of low, average, and high intelligence. Unpublished doctoral thesis, Univer. Wisconsin, 1958.

McGeoch, J. A., & Irion, A. I. *The psychology of human learning.* New York: Longmans, 1952.

Received October 27, 1958.

MICROFORM

All journals published by the American Psychological Association are being made available on MICROFILM or MICROCARD.

Psychological Review
American Psychologist
Psychological Bulletin
Psychological Abstracts
Contemporary Psychology
Psychological Monographs
Journal of Applied Psychology
Journal of Consulting Psychology
Journal of Educational Psychology
Journal of Experimental Psychology
Journal of Abnormal and Social Psychology
Journal of Comparative and Physiological Psychology

Available only in volume units; no single issues.

For MICROFILM, order from:

UNIVERSITY MICROFILMS, INC.
313 North First Street
Ann Arbor
Michigan

For MICROCARD, order from:

J. S. CANNER & COMPANY, INC.
Microcard Division
618 Parker Street
Roxbury 20, Massachusetts

Psychology and Mental Health

A report of the Institute on Education and Training for Psychological Contributions to Mental Health, held at Stanford University in August, 1955.

Edited by CHARLES R. STROTHER

Topics discussed at the Institute:

Training Needs of Psychologists in Community Mental Health

Specialization in Training

Practicum Training

Training for Therapy

Training for Research in the Mental Health Field

Problems of Departmental Organization

Price: $1.75

Order from:

American Psychological Association, Inc.
1333 Sixteenth Street, N. W.
Washington 6, D. C.

GRADUATE EDUCATION IN PSYCHOLOGY

Report of the Conference on Graduate Education in Psychology, sponsored by the Education and Training Board of the American Psychological Association and supported by a grant from the National Institute of Mental Health, U. S. Public Health Service; held at Miami Beach, Florida, November 29 to December 7, 1958.

Prepared by the Editorial Committee:

Anne Roe, *Chairman*,

and

John W. Gustad, Bruce V. Moore,

Sherman Ross, and Marie Skodak

Price $1.50

THE AMERICAN PSYCHOLOGICAL ASSOCIATION
Dept. Grad
1333 Sixteenth Street, N.W.
Washington 6, D. C.

The
Journal
of
Educational
Psychology